LEVEL II

(2)

(6) # AN APPLICATION OF SUBJECTIVE PROBABILITIES TO THE PROBLEM OF UNCERTAINTY IN COST ANALYSIS

(10) HARLEY R./JORDAN
MICHAEL R./KLEIN

(11) NOV[redacted]75

(12) 71

OFFICE OF THE CHIEF OF NAVAL OPERATIONS
Resource Analysis Group
Systems Analysis Division (OP-96D)
Room 2C340 - Pentagon
Washington, DC 20350

406411

ABSTRACT

All cost estimates are characterized by some uncertainty. A device helpful in communicating this uncertainty to the decision maker is a subjective probability distribution of the system cost. A technique--termed the Subjective Probability Estimation Technique (SPET)--is described and a computer program is presented to facilitate its use. This technique permits the analyst to represent his notions about cost uncertainty with the beta or other statistical distributions.

ACKNOWLEDGMENTS

The authors are grateful to Mr. Joseph T. Kammerer, Director of the Resource Analysis Group, for the suggestion and encouragement to pursue this research. Along with many valuable comments, Mr. Kammerer provided the algorithm[1/] for computer program PARAM, which appears in Appendix A of this paper.

Several other individuals offered helpful suggestions during the course of this research, particularly Mr. Carl Wilbourn of the Resource Analysis Group and Mr. Kenneth Linder of the Office of Program Analysis and Evaluation, in the Department of Health, Education, and Welfare.

Those familiar with the literature on uncertainty in cost analysis will recognize a basic similarity between the technique described in this paper and the one expounded by Dienemann.[2/] This paper is a refinement and expansion of certain elements of his earlier research.

1/ See J. T. Kammerer, ASW Force Level Study - Equipment Readiness: Models, Computer Simulation and Results (Washington, DC: Office of the Chief of Naval Operations, 1968).

2/ Paul F. Dienemann, Estimating Uncertainty Using Monte Carlo Techniques (Santa Monica, CA: The RAND Corp., RM-4854-PR, 1966).

CONTENTS

INTRODUCTION

The cost analyst is faced with many uncertainties as he attempts to estimate the costs associated with a new, undeveloped system. He may wonder:

(1) Will the physical characteristics of the system remain unchanged as the development process proceeds?

(2) Will there be any unforeseen problems in the development process?

(3) Will the economic state of the firms or industry responsible for system development and production continue to change as forecasted?

(4) Is the quality of the historical cost data sufficiently high to inspire confidence in the estimates made with it?

(5) Have the cost-estimating relationships been properly specified?

To the extent the analyst is unable to obtain complete answers to these and similar questions, his cost estimates will be enshrouded with uncertainty. Since it is impossible to obtain definitive answers to all these questions, his cost estimates will always be characterized by some uncertainty.

The analyst can treat this uncertainty in one of several ways. He can choose to ignore it and simply report to the decision maker the estimate which represents the "most likely" or "best" estimate of cost, as in the case of this hypothetical guided missile system:

FIGURE 1

"BEST" UNIT COST ESTIMATE OF A HYPOTHETICAL GUIDED MISSILE SYSTEM

This approach, however, belies the existence of a range of possible costs; when the system is finally acquired, any one of the innumerable costs within this range may have been realized. Such an oversimplification may mislead the decision maker by causing him to place excessive confidence in the best cost estimate. An illustration of this potential pitfall is provided in the following example:

> Suppose two alternative systems that do the same task are being compared. Suppose, too, that on balance, the differences in effectiveness, performance, growth potential, maintainability and similar considerations between the two systems are small, so that the choice is primarily one of cost. Suppose one system is estimated to cost $1.25 million and the other $1.50 million. Without an indication of the possible high and low values, the $1.25 million alternative would be the logical choice. But the choice becomes more difficult if, as shown in the accompanying tabulation, the $1.25 million cost is qualified with an estimate of a possible high value of $2.00 million, whereas for the other alternative the limits are estimated to be tighter, and the possible high value is only $1.60 million. The case for

System	Cost Estimate (Millions of Dollars)		
	Lowest	Most Likely	Highest
First Alternative	1.00	1.25	2.00
Second Alternative	1.40	1.50	1.60

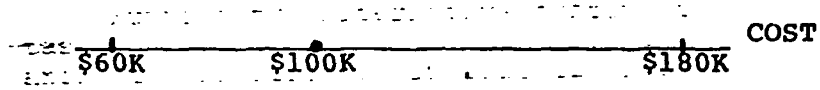
> the alternative with the "most likely" cost of $1.25 million is now more dubious, because its uncertainty spread is greater, extending on the high end to greater costs than the other alternative.3/

The analyst can avoid the problem associated with a single best cost estimate by supplying the decision maker with estimates of the lowest and highest possible costs in addition to the best estimates:

3/ W. Sutherland, Fundamentals of Cost Uncertainty Analysis (McLean, VA: Research Analysis Corp., RAC-CR-4 1971), pp 3-4.

FIGURE 2

"BEST" AND RANGE ESTIMATE OF THE UNIT COST
OF A HYPOTHETICAL GUIDED MISSILE SYSTEM

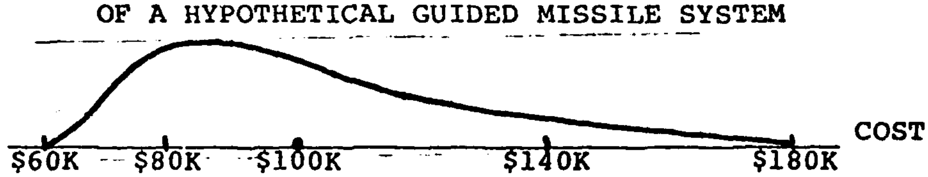


This approach has merit in that it reflects the range of the cost uncertainty. However, it gives little information about the nature of the uncertainty, e.g., whether all the values in the range are almost equally likely to occur, or whether the values closer to the best estimate are much more likely to occur than those near the extremities. Such knowledge could be valuable to the decision maker.

One device helpful in communicating knowledge of both the range and nature of the uncertainty is the probability distribution of the total system cost. To illustrate, consider a probability distribution of the cost of the hypothetical guided missile system:

FIGURE 3

PROBABILITY DISTRIBUTION OF THE UNIT COST
OF A HYPOTHETICAL GUIDED MISSILE SYSTEM

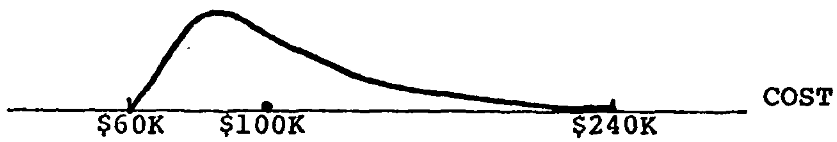


Note that the range of, say, a 95 percent confidence interval [$80K, $140K] is significantly smaller than the full range [$60K, $180K]. The knowledge that the analyst is 95 percent confident that the cost will occur in the much smaller interval [$80K, $140K] will permit the decision maker to act with a more precise idea of the probable cost of the system than he could otherwise (providing, of course, that he trusts the analyst's judgement).

A popular source for the probability distribution of the cost of a weapon system is the prediction interval obtained from cost-estimating relationships (CERs) developed by regression analysis. Unfortunately, this method of probability analysis has a serious limitation: there is no provision for the analyst to incorporate into the analysis his notions about the stochastic behavior of the system cost. For example, most cost analysts have a good idea of the lower bound on the cost, but are less certain about the upper bound. This suggests a probability distribution that is positively skewed:

FIGURE 4

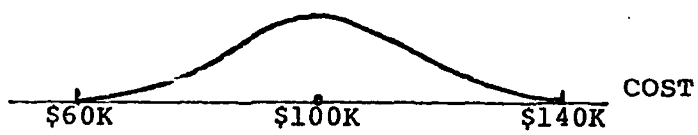
POSITIVELY SKEWED PROBABILITY DISTRIBUTION



The user of the prediction interval obtained from classical regression analysis, however, has to accept a symmetric probability distribution -- the normal probability distribution -- often against his better judgement:

FIGURE 5

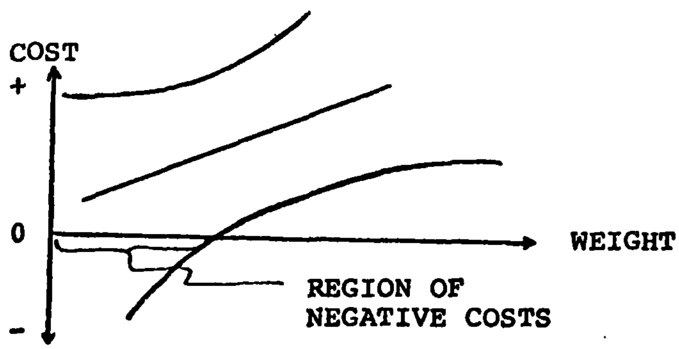
SYMMETRIC PROBABILITY DISTRIBUTION



In addition to this limitation, the paucity (and variability) of data used in regression analysis of weapon system costs often results in prediction intervals with lower bounds which include an extensive region of negative costs:

FIGURE 6

PREDICTION INTERVAL ABOUT A
HYPOTHETICAL REGRESSION OF COST ON WEIGHT



The fact that this model predicts the impossible -- negative costs -- again reveals the above-stated limitation of this approach to probability analysis. What the analyst needs is a technique that will permit him to retain the "most-likely" estimate of system cost and incorporate his a priori beliefs into the prediction interval.

This paper describes such a technique. The authors have entitled it the Subjective Probability Estimation Technique (SPET). This technique is based on the same principles used by Program Evaluation Review Technique (PERT) analysts years ago to treat time-estimating uncertainty.[4] As its name suggests, SPET accounts for the fact that the probability distribution of the cost of a new system is by necessity subjective since repeated observations on the cost of the system, from which an objective probability distribution could be inferred, cannot be made (there is only one observation on the cost of a new system -- the final acquisition cost of the system -- and when this observation is made the need for an estimate terminates). The analyst implements SPET in three steps by:

(1) decomposing the system under examination into several subsystems whose costs are additive;

(2) selecting the subjective probability distribution that best represents his knowledge and judgement about the cost of a subsystem;

(3) combining the subjective probability distribution of each subsystem cost into a subjective probability distribution of total system cost.

The remainder of this paper discusses the theoretical and practical aspects of these three steps.

SYSTEM DECOMPOSITION

When a cost analyst estimates the cost of a complex system he generally breaks the system down into several

---

4/ F. S. Hillier and G. J. Lieberman, *Introduction to Operations Research* (San Francisco, CA: Holden-Day, Inc. 1967), pp. 227 229-232. See also K. R. MacCrimmon and C. A. Ryavec, *An Analytical Study of the PERT Assumptions* (Santa Monica, CA: The RAND Corp. RM-3408-PR 1962).

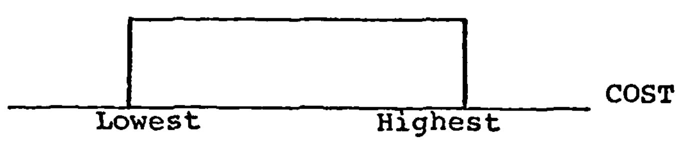
subsystems and estimates the cost of each subsystem. The breakdown of the system is usually determined by the analyst's knowledge of the system and the form of his data base. Using the technique described in this paper, the analyst will also develop a subjective probability distribution describing his uncertainty as to the cost of these subsystems. Of course, if some subsystem cost is known precisely, no uncertainty is involved and this cost is treated as a constant.

SELECTING THE SUBJECTIVE PROBABILITY DISTRIBUTION

A probability distribution can be selected to represent any imaginable combination of factual knowledge and subjective notions an analyst might have about the cost of a subsystem. For example, suppose the analyst has a good idea of what the lowest and highest possible costs for a subsystem could be, but he feels that all costs within that range are equally likely. His subjective probability distribution for the cost of this subsystem can be represented quite adequately with the uniform probability density function:

FIGURE 7

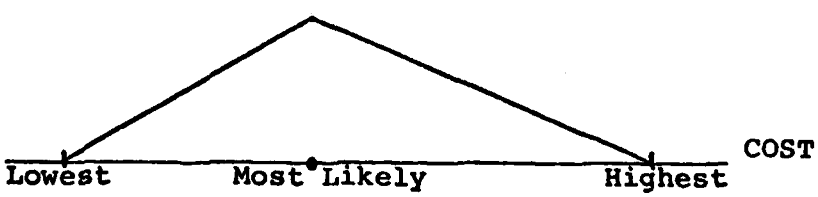
UNIFORM PROBABILITY DENSITY FUNCTION



Suppose that instead of feeling that all costs within the range are equally likely the analyst feels that a particular cost within the range is more likely to be realized than any other. He could represent the cost with a triangular distribution:

FIGURE 8

TRIANGULAR PROBABILITY DISTRIBUTION



or the beta distribution:

FIGURE 9

BETA PROBABILITY DISTRIBUTION

or any other distribution with a single maximum value.

The beta distribution is one of the most popular among cost analysts for representing subjective probability distributions. The popularity of the beta is due to its several appealing characteristics. One of these characteristics is that its range may be restricted to positive values; costs similarly are positive. Another is that the beta has a finite, rather than infinite range; it is reasonable to suppose that the cost is bounded by finite upper and lower bounds. Finally, the beta distribution can be expressed in an infinite variety of skewed and symmetric forms which provide the analyst considerable choice when specifying the particular shape of the distribution.5/

Because of the popularity of the beta distribution, the discussion in the remainder of this section will center on it. In the next section, however, a computer program is described that permits the analyst to use any imaginable subjective probability distribution to represent subsystem cost.

5/ *Another commonly used distribution is the Weibull. Most of the appealing properties of the beta are also found in the Weibull distribution. For examples of the use of the Weibull in treating uncertainty in cost analysis see D. F. Schaefer, et. al., A Monte Carlo Simulation Approach to Cost-Uncertainty Analysis, (McLean, VA: Research Analysis Corp.; RAC-TP-349, 1969) and W. H. Sutherland, A Method for Combining Asymmetric Three-Value Predictions of Time or Cost (McLean, VA: Research Analysis Corp.; RAC-P-65, 1972).*

The usual expression for the beta probability density function (pdf) is:6/

$$f(x) = Cx^a(1 - x)^b; \; 0 < x < 1; \; a,b > 0; \qquad [1]$$

where $C = \Gamma(a + b + 2)/[\Gamma(a + 1)\Gamma(b + 1)]$

= the inverse of the complete beta function

and $\Gamma(t) = \int_0^\infty z^{t-1}e^{-z}dz, \; t > 0.$

This version of the beta pdf will be called the normalized beta pdf, since the range of x is the unit interval.

A simple linear transformation, $x^* = L + (H - L)x$, where L and H are the lowest and highest values of x^*, respectively, extends the range of x in equation [1] to any finite interval, yielding a generalized beta pdf:

$$g(x^*) = [C/(H - L)^{a+b+1}](x^* - L)^a(H - x^*)^b \qquad [2]$$

where C is as defined in equation [1] and $L < x^* < H$, $a,b > 0$.

The four parameters of the generalized beta pdf are a, b, L, and H.7/ Therefore, a unique pdf is defined for every four-tuple (a,b,L,H). The values the analyst assigns to these parameters can be obtained through certain estimation procedures.

The analyst may estimate L and H directly from his and other expert knowledge of the subsystem's technology, contractor (builder), industry, etc. After analyzing

6/ H. J. Larson, *Introduction to Probability Theory and Statistical Inference* (New York: John Wiley and Sons, Inc., 1969), p. 305; B. W. Lindgren, *Statistical Theory* (New York: The MacMillian Co., 1968), p. 373.

7/ Note that L and H serve only to specify the origin and range of x^*, whereas a and b determine the shape of the pdf of x^*.

these data, he chooses L and H so that the cost of the subsystem could never be less than L or greater than H.

Estimates of the parameters a and b, however, cannot be obtained in a direct manner. One way to estimate them is to obtain two functionally independent equations in a and b and then solve them for these parameters. In Appendix A, two such equations are proposed and a computer program facilitating their solution is documented. Potential users of this method for estimating a and b are cautioned; some combinations of the analyst-supplied inputs result in distributions that cannot be represented by a beta random variable. This problem can be avoided by using another pair of equations. Consider the mode $M(x^*)$ and the variance $V(x^*)$ of the generalized beta distribution:8/

$$M(x^*) = (aH + bL)/(a + b), \quad L \leqq M(x^*) \leqq H \qquad [3]$$

$$V(x^*) = [(a + 1)/(b + 1)(H - L)^2]/[a + b + 2)^2$$

$$(a + b + 3)], \quad 0 \leqq V(x^*) \leqq (H - L)^2/12 \qquad [4]$$

Using cost-estimating relationships or other methods, the analyst can estimate the mode or most likely value of the subsystem cost.9/ By evaluating much of the information he has about the subsystem cost he can estimate its

---

8/ See Footnote 2 in Appendix A for the derivation of these formulae.

The range of the mode is obtained from its definition. The range of the variance is due, in part, to the fact that the lower bound on the variance of any distribution is zero, and, in part, to the fact that the beta distribution converges to the uniform as parameters a and b approach zero. The upper bound on the variance of the generalized beta distribution is, therefore, the variance of the (generalized) uniform distribution, namely

$$V(x^*) = (H - L)^2/12 \qquad [5]$$

9/ Under most circumstances the most-likely value is the analyst's point estimate of the subsystem cost.

variance.[10] Having obtained estimates for these two statistics, he can solve equations [3] and [4] simultaneously to determine unique values (estimates) for a and b.

The principal difficulty with the technique proposed is in the estimation of the variance. It is difficult for the analyst to interpret his beliefs concerning the uncertainty surrounding a point estimate in terms of the beta variance. As an aid in this process, the analyst may consider a related variable, which is termed an uncertainty coefficient in this paper. The uncertainty coefficient represented with the letter "U" is a normed linear measure of the analyst's uncertainty. If there is no uncertainty in the estimate, $U = 0$; if there is total uncertainty on the whole range (i.e., all values are equally likely), $U = 1$. In most instances, the analyst can assign a reasonable value to U. The variance of x^* can then be determined from the relationship:

$$V(x^*) = [(H - L)U]^2/12, \quad 0 < U < 1. \qquad [6]$$

It is difficult for the analyst to visualize the distribution he has chosen from his estimates of the parameters. However, since the shape of the beta distribution is determined by two of the parameters (a and b) which are in turn determined by the values of the mode and the uncertainty coefficient, it is possible to get a reasonable

---

10/ *It has been proposed that the analyst assume that the range of the cost variable is equal to six standard deviations, yielding $V(x^*) = [(H - L)/6]^2$. The basis for this assumption is that "most" of the probability associated with distributions such as the normal distribution is contained in the interval ± 3 standard deviations from the mean. The authors of this paper do not find this to be a very strong motivation. The total range of the uniform pdf is contained in an interval of ± 1.75 standard deviations and this distribution is a limiting form of the beta distribution. Further, the significance of the behavior of infinite, symmetric distributions such as the normal pdf in deriving properties of the (generally) finite, asymmetric beta distribution is questionable. It seems more logical to allow the analyst to input his knowledge of the uncertainty via the uncertainty coefficient. However, if the user prefers this device he should input the value U = 0.577 when using program SPET.*

idea of the distributional form from a set of normalized graphs of beta pdfs with different modes and uncertainty coefficients. Such a set appears in Appendix B.

Appendix B contains ten groups of graphs of normalized beta pdfs. Each group contains graphs of three pdfs with the same mode but distinct uncertainty coefficients. The modal values vary from set to set, beginning with .05 and increasing by .05 until .50 when reading the abscissa from left to right, or beginning with .95 and decreasing by .05 until .50 when reading from right to left. To use Appendix B, the analyst computes the estimated normalized mode (N) from his estimates of L, H, and M with the relationship

$$N = (M - L)/(H - L) \qquad [7]$$

and then selects the group of pdfs whose normalized mode is closest to this computed N. From the three graphs in this subset the analyst can see how the pdf varies with the uncertainty coefficient and get a reasonable idea of the shape of the distribution he has chosen. Alternatively, he may look at the set before choosing the uncertainty coefficient and use the information he gains to help him select the value for U.

<u>Example</u>

Assume an analyst is studying the cost of some subsystem "S". By analogy with other systems, or by some other technique, he determines that the cost of S will be something greater than $7,000 but less than $12,000. Further, utilizing a CER, or by some other technique, he estimates its most likely value at $10,500. He calculates the normalized mode "N" from the equation:

$$N = \frac{M - L}{H - L} = \frac{10,500 - 7,000}{12,500 - 7,000} = .64$$

The set of graphs corresponding to this system is found between pages B-19 and B-21. These graphs are read from right to left.

After the analyst has developed the distributions of each subsystem he faces the problem of determining the distribution of their sum (the total cost of the system).

COMBINING SUBJECTIVE PROBABILITY DISTRIBUTIONS

Of the several techniques that have been employed by cost analysts to combine statistical distributions representing their subjective probability distributions, two of the more popular are derivation of moments11/ and Monte Carlo simulation.12/ Each of these techniques has its advantages: the former can be done with tables and a desk calculator, whereas the latter requires access to an electronic computer. The latter, however, is much faster and easier to use. For this reason Monte Carlo simulation is the technique employed in this research. Appendix C documents Program SPET, a computer program for adding independent statistical distributions by means of Monte Carlo simulation (SPET also performs other calculations discussed in the next section).

Program SPET has been used successfully on an interactive time-sharing computer system. Basically the user enters the parameters of the statistical distributions selected by the analyst and the program generates frequency distributions and summary statistics of the total system cost. Details and an example of the inputs, outputs, and operation of the program can be found in Appendix C.

THE INDEPENDENCE ASSUMPTION: A PROBLEM

When the analyst decomposes the system he is costing into several subsystems, it is very unlikely that the costs of the various subsystems are always statistically independent of one another. For example, the cost of the propulsion system of a guided missile is probably correlated with the cost of its payload. The correlation would be positive if an increase in payload cost was due to an increase in payload size which, in turn, would require a more powerful and hence more costly propulsion system. On the other hand, the correlation would be negative if an increase in payload cost was due to a reduction in payload

---

11/ See S. Sobel, *A Computerized Technique to Express Uncertainty in Advanced System Cost Estimates* (Bedford, MA: The Mitre Corp., TM-3728, 1963), and W. H. Sutherland, *A Method for Combining Asymmetric Three-Value Predictions of Time or Cost* (McLean, VA: Research Analysis Corp., RAC-P-65, 1972).

12/ P. F. Dienemann, *op. cit*, and D. F. Schoefer, *op. cit*.

size brought about by miniaturization which, in turn, would require a less powerful and hence less costly propulsion system. It is impossible to determine a priori whether the correlation between the costs of any two subsystems is positive or negative, but the experience of the authors suggests that in the majority of cases it will be positive.

If an analyst assumes that the statistical distributions (random variables) representing subsystem costs are independent when in fact they are positively correlated, he will underestimate the variance of their sum. To see this, consider the expression for the sum (S) of the variance of n random variables (X_i):

$$V(S) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov(X_i, X_j) \qquad [8]$$

The assumption of independence implies that $Cov(X_i, X_j) = 0$ for all $i \neq j$, which in turn implies that the expression

$$2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov(X_i, X_j) = 0.$$

If the X_i are dependent, this term could be positive, negative, or zero. If it is positive, its deletion from [8] by assuming independence results in an understatement of V(S). An understatement of V(S) could be a serious problem because it results in a confidence interval about the mean of the total system cost distribution that is smaller than it should be. This might cause the decision maker to posit unwarranted confidence in the estimate.

Assuming the random variables representing subsystem costs are positively correlated, the magnitude of the underestimate of V(S) is directly related to the number of variables in the sum. To illustrate this fact, consider the following example:

There are two positively correlated random variables, X_1 and X_2. Then

$$V(S) = V(X_1) + V(X_2) + 2Cov(X_1,X_2) \qquad [9]$$

Now assume $X_1 = Z_1 + Z_2$. Then

$$V(S) = V(Z_1) + V(Z_2) + V(X_2) + 2\ \mathrm{Cov}(Z_1,Z_2) + 2\ \mathrm{Cov}(Z_1,X_2) + 2\ \mathrm{Cov}(Z_2,X_2)$$

$$= V(Z_1) + V(Z_2) + V(X_2) + 2\ \mathrm{Cov}(X_1,X_2) + 2\ \mathrm{Cov}(Z_1,Z_2) \quad [10]$$

Assuming independence in the case of two variables results in the deletion of 2 $\mathrm{Cov}(X_1,X_2)$ from V(S). However, in the case of three variables the assumption of independence results in the deletion of not only 2 $\mathrm{Cov}(X_1,X_2)$ but 2 $\mathrm{Cov}(Z_1,X_2)$ as well. Clearly, if the covariances are positive, V(S) is understated more in the case of three variables than in the case of two.

Therefore in the case where the random variables representing subsystem costs are positively correlated not only is the variance of the total system cost underestimated, but the magnitude of the underestimate is directly related to the number of variables making up the sum.

The obvious solution to this problem is to incorporate into computer Program SPET provisions for the consideration of probable correlation among the variables and then proceed to estimate the nature of the correlation. Although the former idea presents no problem, the latter appears to be a most difficult task. It is not clear at this point how to systematically estimate the correlation among the variables representing subsystem costs. Hopefully a credible technique for doing such will become apparent to someone.

Program SPET has been designed to provide some insight into the significance of the independence assumption in two ways. First, by using the same random number in sampling from all the subsystem distributions, a distribution of total cost which the printout titles "Dependent Beta" is derived. The technique imposes a functional relationship between all the variables. Note that this functional relationship is not an arbitrary relationship but is imposed by the forms of the subsystem pdfs developed by the analyst. Specifically, if $F_i(X_i)$ is the cumulative distribution function of the ith subsystem then:

$$\sum_{i=1}^{n} X_i = X_1 + \sum_{i=2}^{n} F_i^{-1}[F_1(X_1)] \quad [11]$$

where F_i^{-1} is the functional inverse of F_i. The functions F_i and F_i^{-1} are much too complex to derive the specific form of the relationship imposed but the technique does impose a very real positive correlation between the variables.

As a second means of examining uncertainty without imposing the independence assumption, SPET prints the statistics for a uniform distribution of total cost on the interval between the sum of the minimum costs of the subsystems and the sum of the maximum cost of the subsystems. This is meant to serve as a "worst case." SPET also prints the total cost distribution assuming each of the subsystem's costs is uniformly and independently distributed.

SUMMARY

Several conceptual points have been discussed, among them:

(1) the nature of uncertainty in cost analysis;

(2) the value of treating uncertainty explicity in cost analysis;

(3) a problem inherent in using the classical linear regression model as a basis for statements on cost uncertainty;

(4) the subjective nature of cost uncertainty;

(5) the properties of the beta distribution and how they can be used to facilitate cost uncertainty analysis; and

(6) the dependence of subsystem costs and its impact on statements about uncertainty.

The basic practical contribution of this paper is a computer program for generating statements on cost uncertainty that permits the analyst to input any imaginable probability distribution to represent a subsystem cost.

APPENDIX A

ESTIMATION OF PARAMETERS a AND b:

COMPUTER PROGRAM PARAM

Introduction

One way an analyst can estimate parameters a and b of the generalized beta probability density function (pdf)[1] is to simultaneously solve two functionally independent equations in a and b. Consider the mode -- the "most likely" value -- of the generalized beta pdf:

$$M(x^*) = (aH + bL)/(a + b)^{2/} \quad [A1]$$

Using a cost-estimating relationship, or other methods, the analyst can estimate the mode (M) of the subsystem cost he wants to represent with a beta pdf. Substitution of M into equation [A1], along with L and H, yields one equation in a and b:

$$M = (aH + bL)/(a + b) \quad [A7]$$

Another equation in a and b can be obtained from an estimate of the probability (P) that the cost of the sub-

---

1/ *See equation [2] on page 8.*

2/ *This expression can be obtained by solving $d[g(x^*)]/dx^* = 0$ for x^*, where $g(x^*)$ is given by equation [2]. However, it is easier to obtain expressions for the mode as well as the mean, $E(x^*)$, and the variance, $V(x^*)$, of the generalized beta pdf by means of simple algebraic operations on the expressions for these statistics derived from the more familiar normalized beta pdf. The mode, mean, and variance of the normalized beta pdf are:*

$$M(x) = a/(a + b) \quad [A2]$$

$$E(x) = (a + 1)/(a + b + 2) \quad [A3]$$

$$V(x) = [(a + 1)(b + 1)]/[(a + b + 2)^2(a + b + 3)] \quad [A4]$$

Note that $M(x) = a/(a + b) = M[(x^* - L)/(H - L)] = [M(x^*) - L]/(H - L)$; solving for $M(x^*)$ yields equation [A1]. Proceeding similarly for $E(x^*)$ and $V(x^*)$ yields

$$E(x^*) = (aH + bL + H + L)/(a + b + 2) \quad [A5]$$

$$V(x^*) = [(a + 1)(b + 1)(H - L)^2]/[(a + b + 2)^2(a + b + 3)] \quad [A6]$$

system will lie within a subinterval of its range [L, H]. For convenience, this subinterval is taken as the interval from L to the midpoint between L and M, i.e., [(L + M)/2]. Then an equation in a and b results from the relationship

$$P = [C/(H - L)^{a+b+1}] \int_{L}^{\frac{L+M}{2}} (x^* - L)^a \; (H - x^*)^b \; dx^* \qquad [A8]$$

where C, x, a, and b are defined in equation [2].

Equations [A7] and [A8] comprise two functionally independent equations in a and b, and when solved simultaneously determine unique values (estimates) for a and b.

Program Description

Program PARAM is written in FORTRAN IV for use in conjunction with a PDP-10 computer in an interactive time-sharing mode. It is designed to solve equations [A7] and [A8] for parameters a and b, given values for L, H, M, and P. This is accomplished in the following sequence:

1 - The inputs L, H, and M are normalized. Denoting the normalized counterparts of L, H, and M by l, h, and m, respectively,

$$l = 0$$
$$h = 1$$
$$m = (M - L)/(H - L)$$

2 - A standard root-finding technique[3/] is employed in Subroutine Beta to find the values for a and b that satisfy

$$P = C \int_{0}^{m/2} x^a (1 - x)^b \; dx \qquad [A9]$$

where P is supplied by the user, m is determined from a user supplied datum (M), and C, x, a, and b are as defined in equation [2]. Note that [A9] is the normalized version of [A8].

3/ *The root-finding technique is known as the "false position" method and is found in most texts on numerical analysis.*

Subroutine Beta calls Function Subprogram Gamma to compute values for $\Gamma(n)$. This is accomplished by use of the relation

$$\Gamma(n+1) = n\Gamma(n) \qquad [A10]$$

and an interpolation procedure.

By repeated application of [A10], $\Gamma(n)$ for any $n>0$ can be expressed as the product

$$(n - 1)(n - 2)\ldots\Gamma(n^*) \qquad [A11]$$

where $1 \leq n^* \leq 2$. Since $\Gamma(n)$ is well-behaved in the range $1 \leq n^* \leq 2$, the values of $\Gamma(n^*)$ can be approximated using a "table look-up" interpolation procedure. Function Subprogram Gamma uses such an interpolation device in conjunction with relation [A11] to evaluate $\Gamma(n)$.

3 - The four-tuple (a,b,L,H) is printed as output.

User Instructions

The following example demonstrates the use of Program PARAM. Note that the user supplied portions of the example are underlined:

```
.RUN PARAM


ESTIMATES OF L,H,M,P
8000,14000,10500,.14


ALPHA IS      1.07
BETA IS       1.50
LOW IS     8000.00
HIGH IS   14000.00


CP UNITS    2

EXIT
```

Program Listing

```
      A=BB
220   PP=0.
      NMIID=0.
      RETURN
50    WRITE(II,60)
60    FORMAT(1H-,28HBETA PARAMETERS OUT OF RANGE)
      RETURN
      END
C
      FUNCTION GAMMA(GP)
      REAL G(22)
      DATA G/1.,.9735,.95135,.93304,.91817,
    + .9064,.89747,.89115,.88726,.88565,
    + .88623,.88887,.89352,.90012,.90864,
    + .91906,.93138,.94561,.96177,.97983,1.,1./
      II=16
      JJ=16
      ERROR=0.
      SUM=1.
      IF(GP-57.4) 10,10,20
10    IF(GP-1.00E-20) 20,30,30
20    WRITE(II,40)
40    FORMAT(1H-,28HGAMMA PARAMETER OUT OF RANGE)
      ERROR=ERROR+1.
      RETURN
30    IF(GP-2.) 50,50,60
60    SUM=SUM*(GP-1.)
      GP=GP-1.
      GO TO 30
50    IF(GP-1.) 70,80,80
70    SUM=SUM/GP
      GP=GP+1.
80    I=(GP-1.)/.05+1.
      XI=I-1
      GPL=1.+XI*.05
      GAMFN=G(I)+(G(I+1)-G(I))*(GP-GPL)/.05
      GAMMA=GAMFN*SUM
      RETURN
      END
```

```
      IF(GP-57.4) 40,40,50
40    C1=GAMMA(GP)
      GP=A+1.
      C2=GAMMA(GP)
      C3=C1/C2
      GP=B+1.
      C4=GAMMA(GP)
      C=C3/C4
70    XMM=XMM+2.
      DO 80 I=2,3
      FT(I)=T(I)**A*(1.-T(I))**3
      T(I+1)=T(I)+SP
80    CONTINUE
      T(2)=T(4)
      XSUM=FT(1)+4.*FT(2)+FT(3)
      FT(1)=FT(3)
      XINTEG=XINTEG+XSUM
      IF(XMM-INTEG) 70,90,90
90    AA=P-C*SP/3.*XINTEG
      IF(ABS(AA)-.0001) 100,100,110
110   IF(AA*AAS) 120,130,130
130   IF(AA) 140,140,150
140   AA1=AA
      GP1=B
      GO TO 160
150   AA2=AA
      GP2=B
160   GO TO (170,180),JJ1
120   JJ1=2
      GO TO 130
170   B=W
      W=W*2.
      GO TO 190
180   B=(AA2*GP1-AA1*GP2)/(AA2-AA1)
190   AAS=AA
      GO TO 30
100   IF(PP-NMIDD) 200,200,210
200   GO TO 220
210   BB=B
      B=A
```

```
          REAL L,M
          II=16
          JJ=16
          WRITE(II,10)
   10     FORMAT(1H-,20HESTIMATES OF L,H,M,P/)
          READ(JJ,20) L,H,M,P
   20     FORMAT(4F)
          CALL BETA(L,H,M,P,A,B)
          WRITE(II,30) A,B,L,H
   30     FORMAT(1H-,9HALPHA IS ,F8.2/10H BETA IS  ,F8.2/
     +    10H LOW IS    ,F8.2/10H HIGH IS  ,F8.2)
          WRITE(II,40)
   40     FORMAT(1H-)
          END
C
          SUBROUTINE BETA(L,H,M,P,A,B)
          REAL L,M,INTEG,NMODE,NMID,NMIDD,FT(3),T(4)
          II=16
          JJ=16
          INTEG=40.
          SEARCH=20.
          NMODE=(M-L)/(H-L)
          NMID=NMODE/2.
          IF(P-NMID) 10,10,20
   20     PP=P
          NMIDD=NMID
          P=1.-P
          NMID=1.-NMID
          NMODE=1.-NMODE
   10     SP=NMID/INTEG
          AAS=-1.
          JJ1=1
          W=SEARCH*(1.-NMODE)
          B=0.
   30     A=B*NMODE/(1.-NMODE)
          XMM=0.
          XINTEG=0.
          T(2)=SP
          FT(1)=0.
          GP=A+B+2.
```

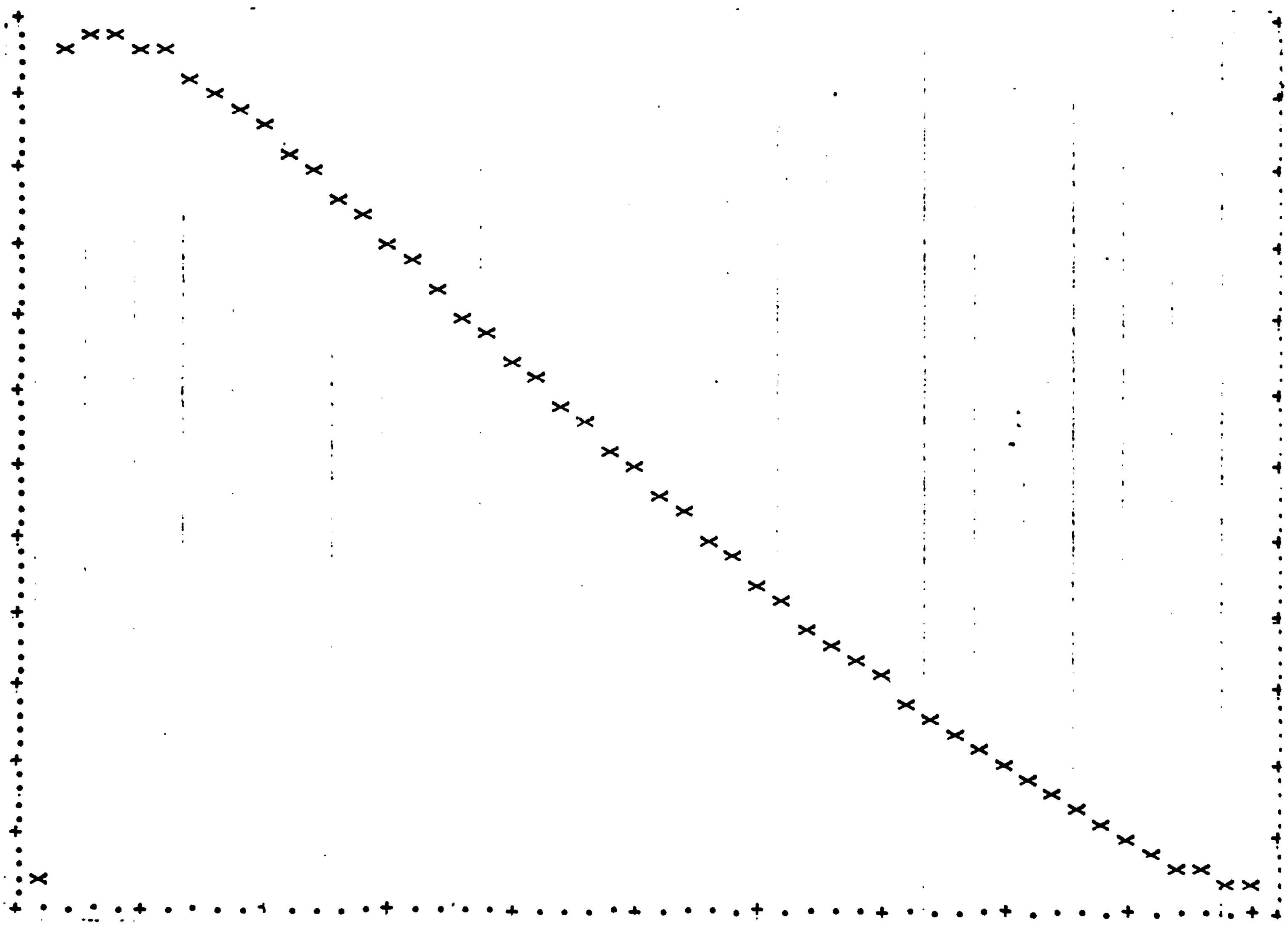
APPENDIX B

GRAPHICAL AIDS FOR SELECTING THE UNCERTAINTY COEFFICIENT

<u>LEFT-TO-RIGHT</u>
Normalized Mode = .05

<u>HIGH VARIANCE</u>
Uncertainty Coefficient = .75

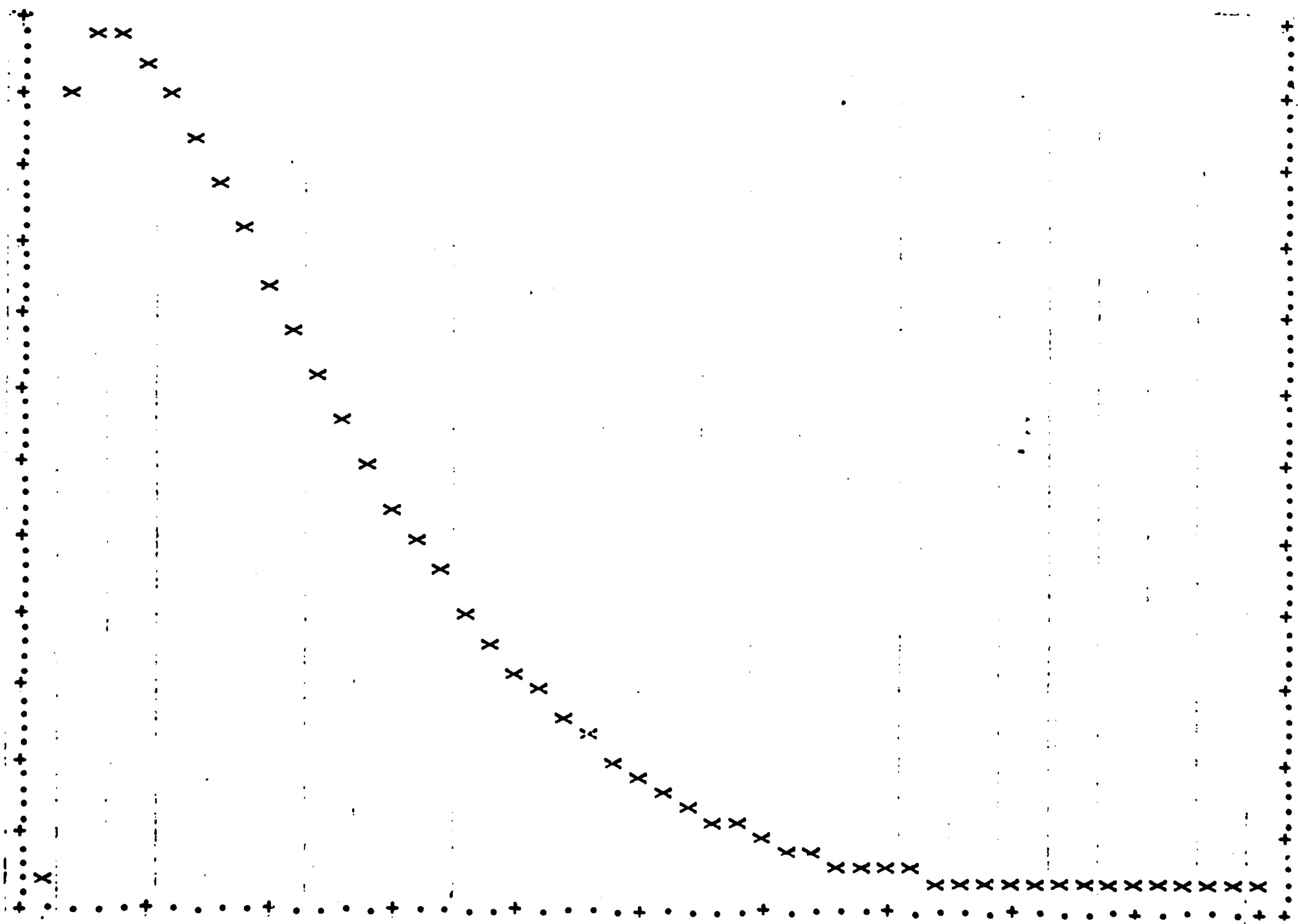
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .95



<u>LEFT-TO-RIGHT</u>
Normalized Mode = .05

<u>MEDIUM VARIANCE</u>
Uncertainty Coefficient = .50

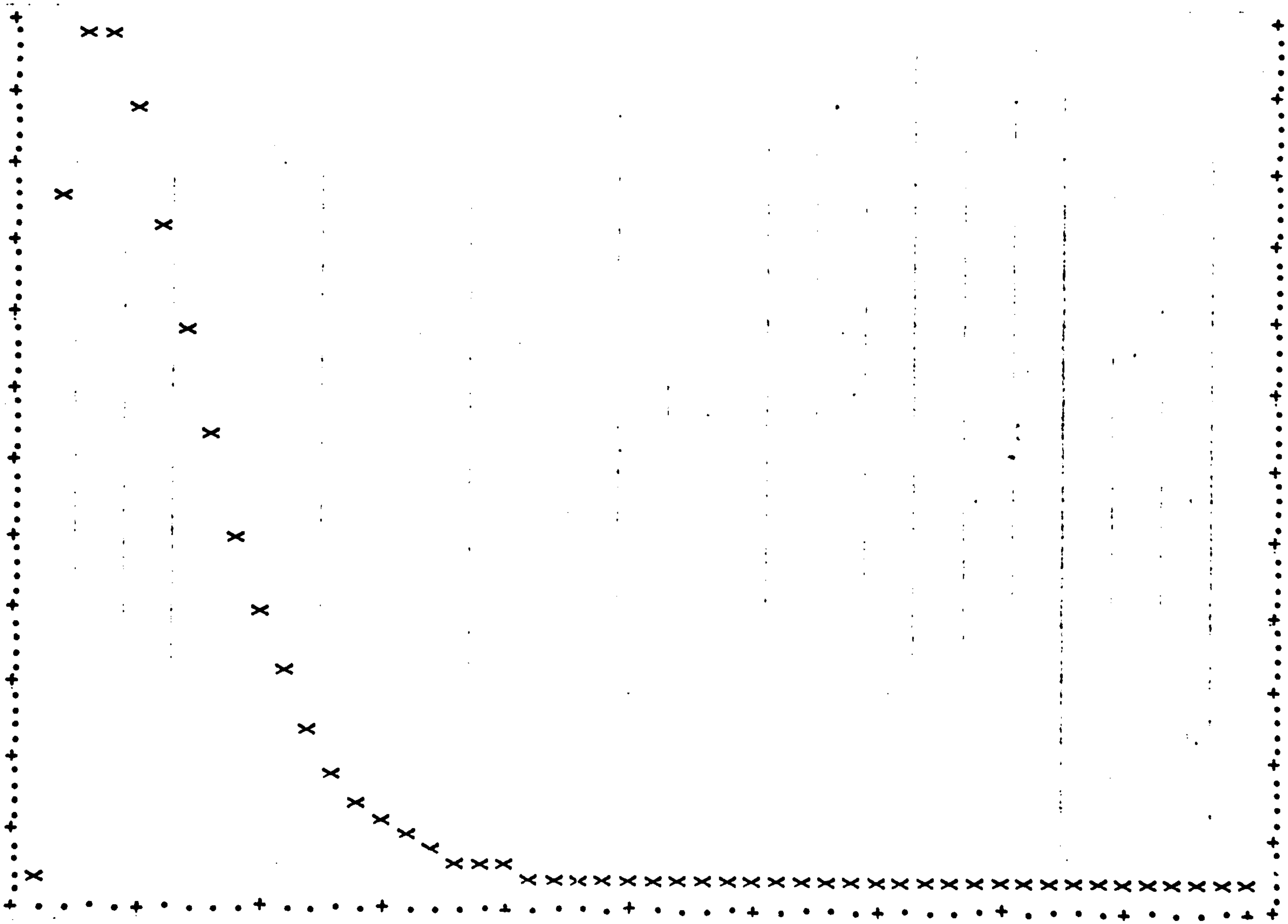
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .95



LEFT-TO-RIGHT
Normalized Mode = .05

LOW VARIANCE
Uncertainty Coefficient = .25

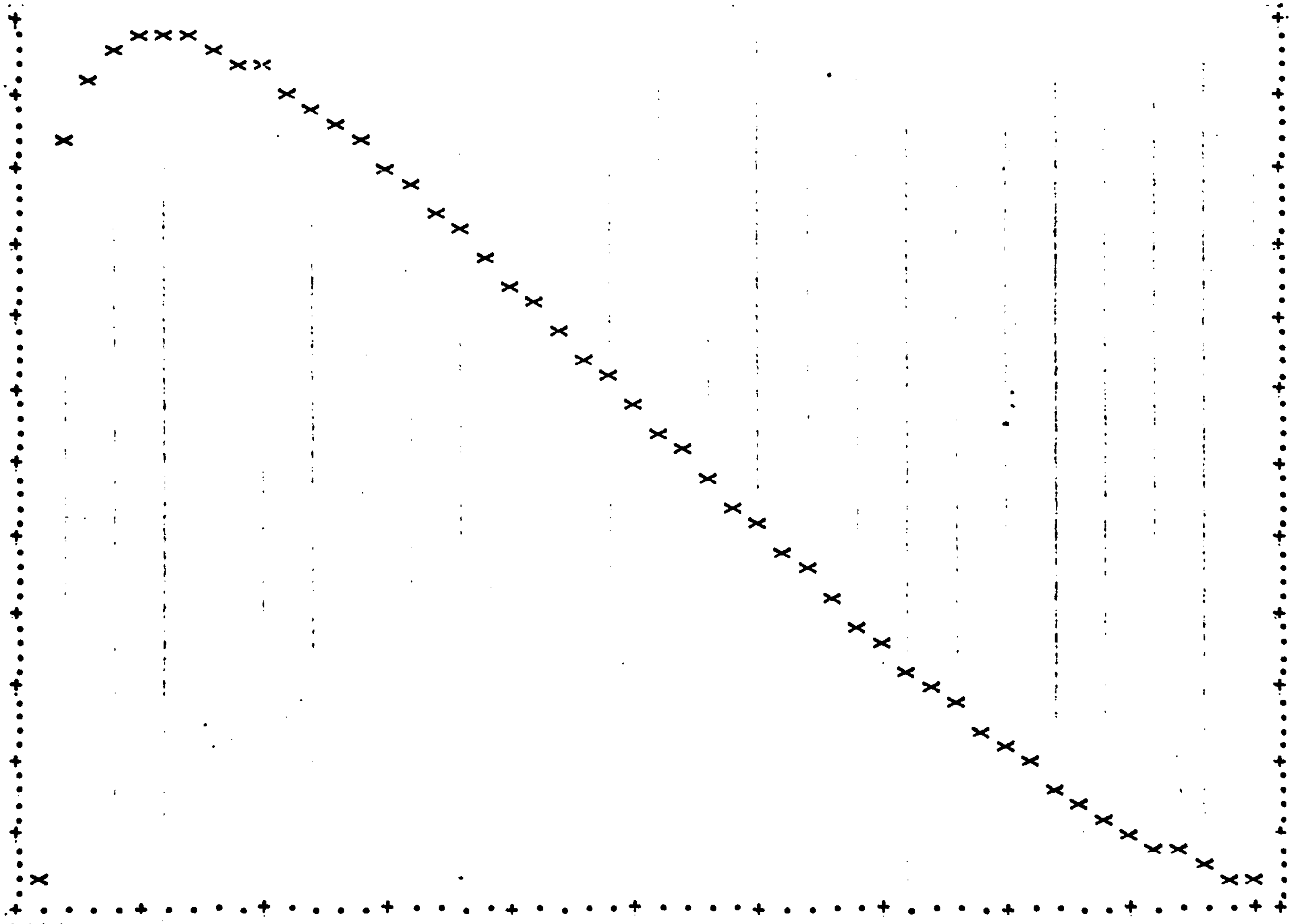
RIGHT-TO-LEFT
Normalized Mode = .95



LEFT-TO-RIGHT
Normalized Mode = .10

HIGH VARIANCE
Uncertainty Coefficient = .75

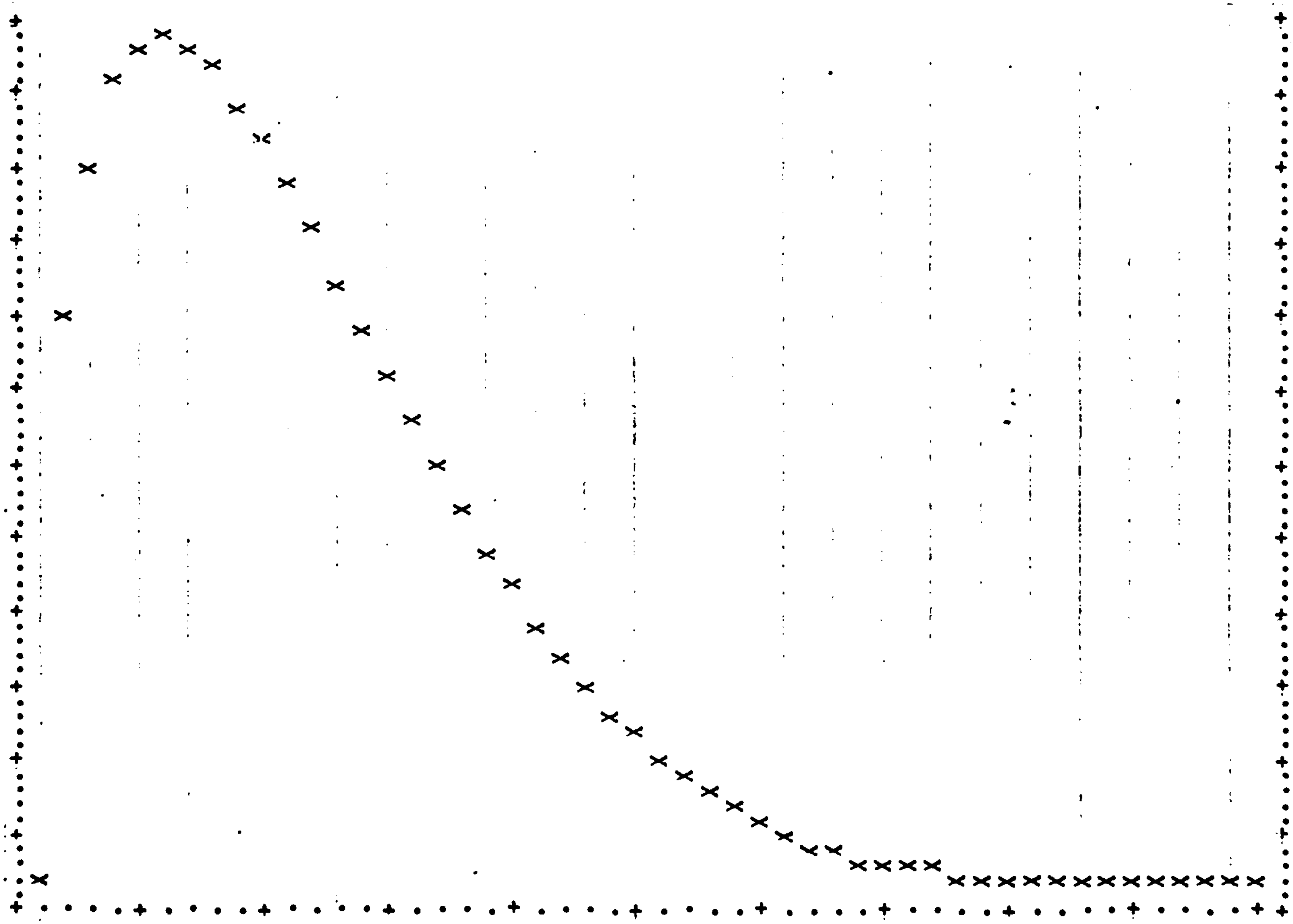
RIGHT-TO-LEFT
Normalized Mode = .90



<u>LEFT-TO-RIGHT</u>
Normalized Mode - .10

<u>MEDIUM VARIANCE</u>
Uncertainty Coefficient = .50

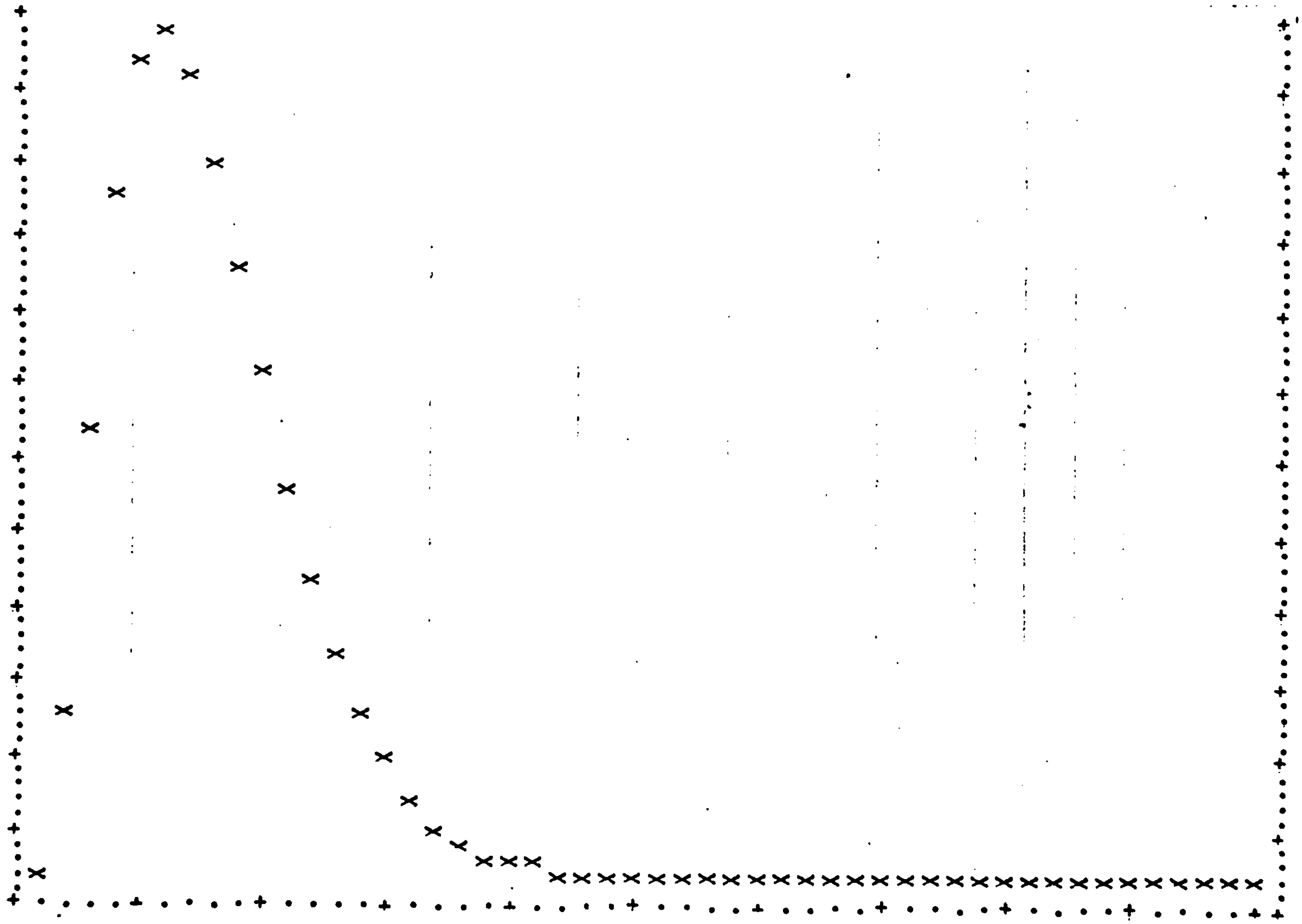
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .90



LEFT-TO-RIGHT
Normalized Mode = .10

LOW VARIANCE
Uncertainty Coefficient = .25

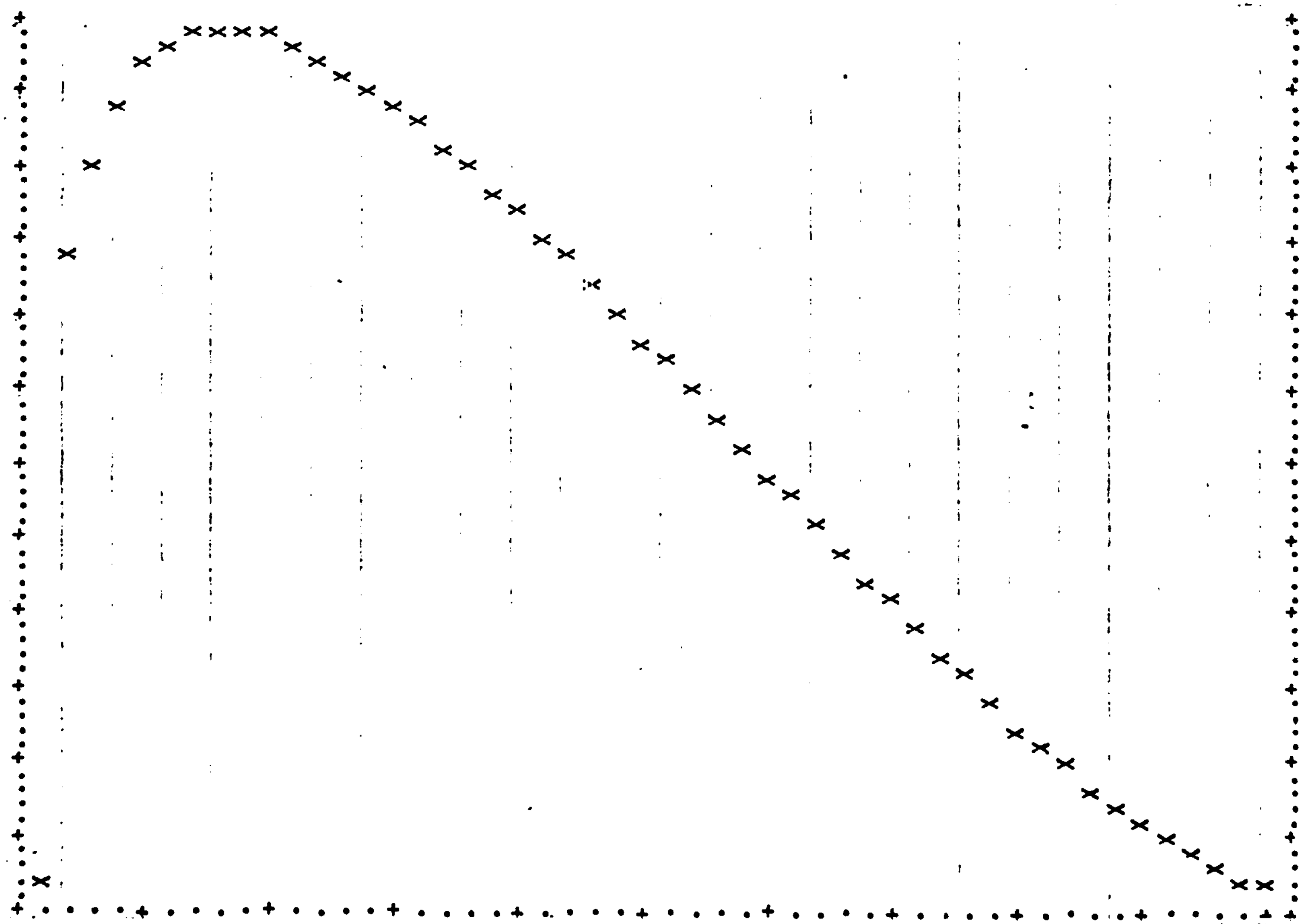
RIGHT-TO-LEFT
Normalized Mode = .90



__LEFT-TO-RIGHT__
Normalized Mode = .15

__HIGH VARIANCE__
Uncertainty Coefficient = .75

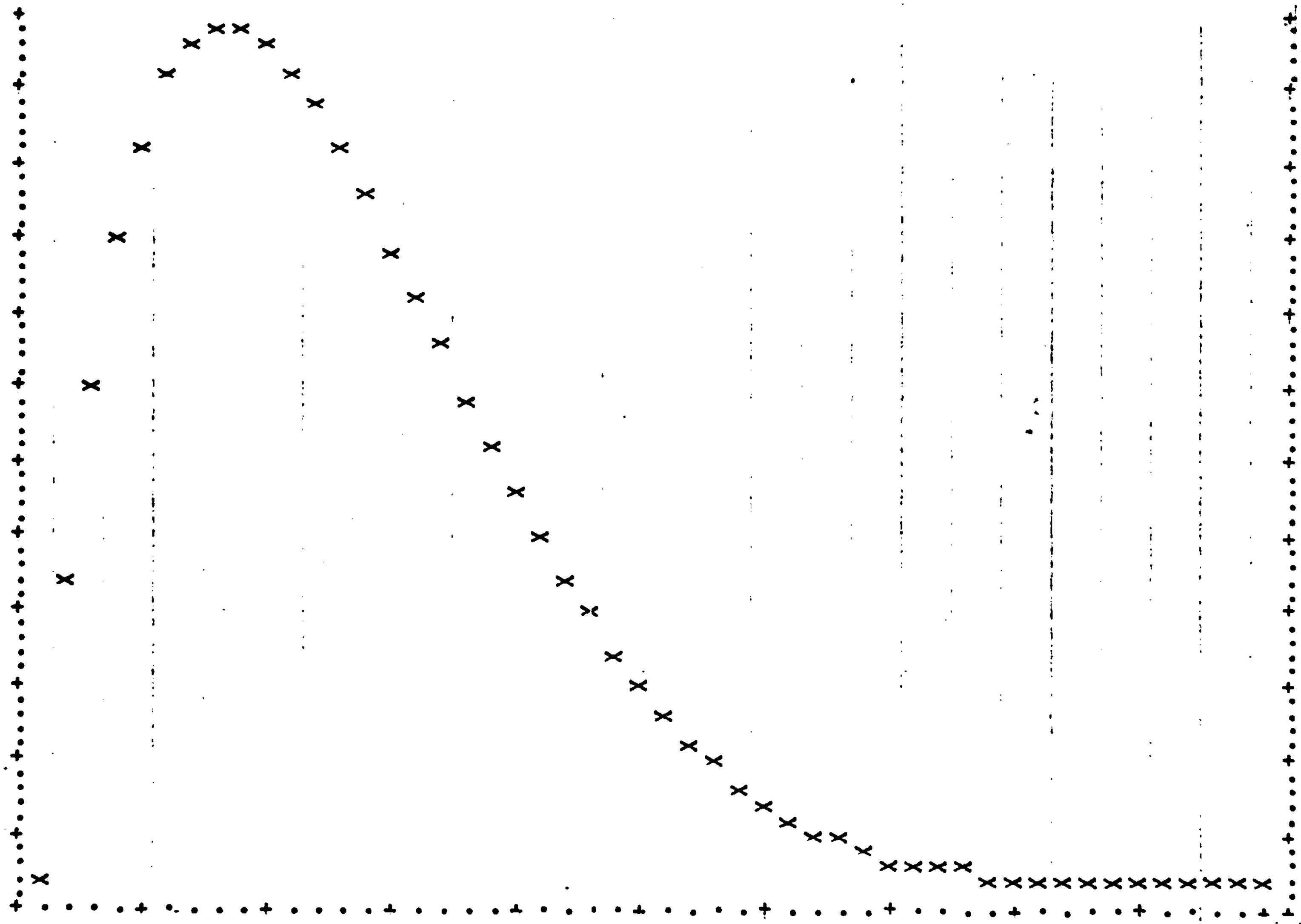
__RIGHT-TO-LEFT__
Normalized Mode = .85



LEFT-TO-RIGHT
Normalized Mode = .15

MEDIUM VARIANCE
Uncertainty Coefficient = .50

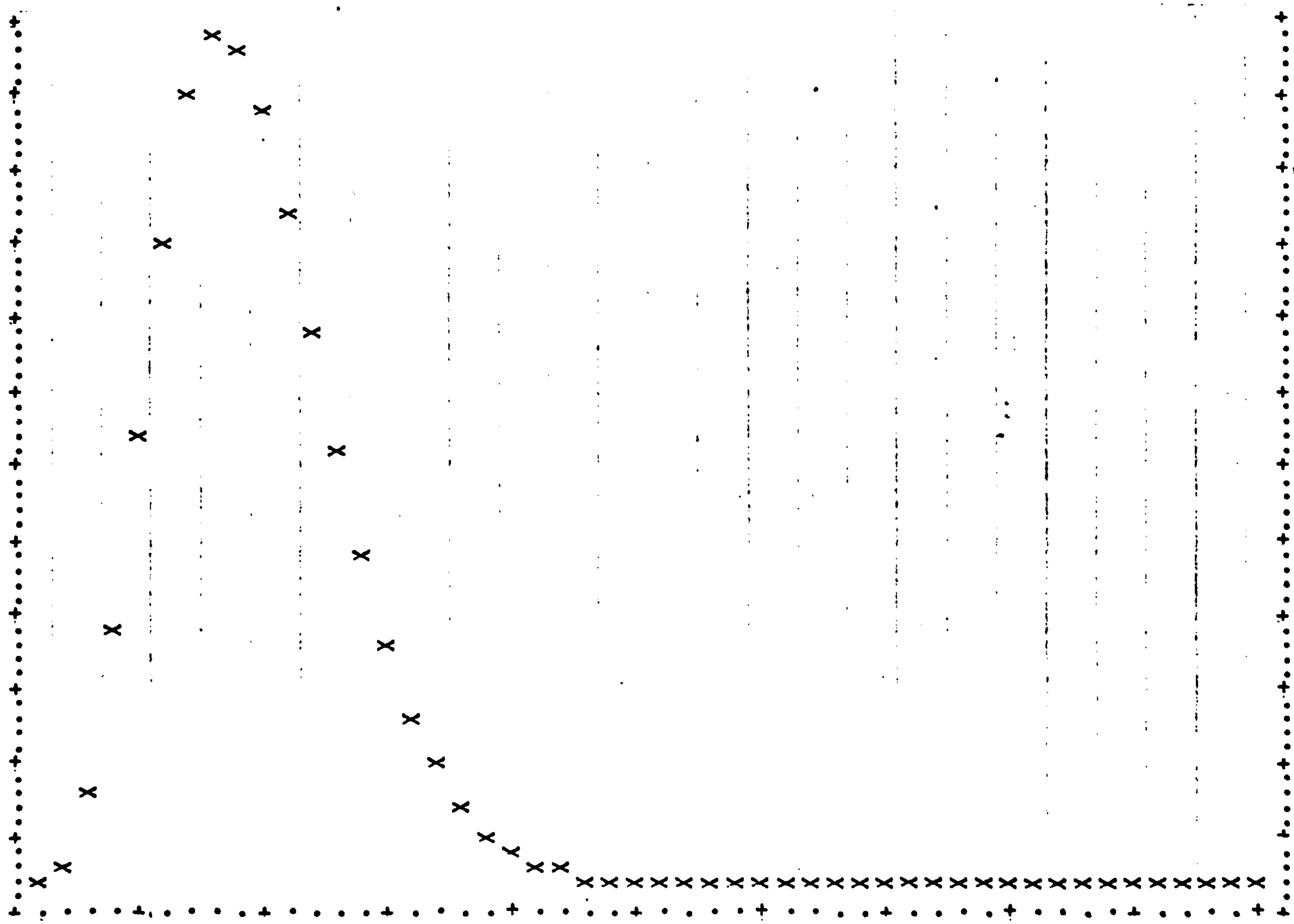
RIGHT-TO-LEFT
Normalized Mode = .85



LEFT-TO-RIGHT
Normalized Mode = .15

LOW VARIANCE
Uncertainty Coefficient = .25

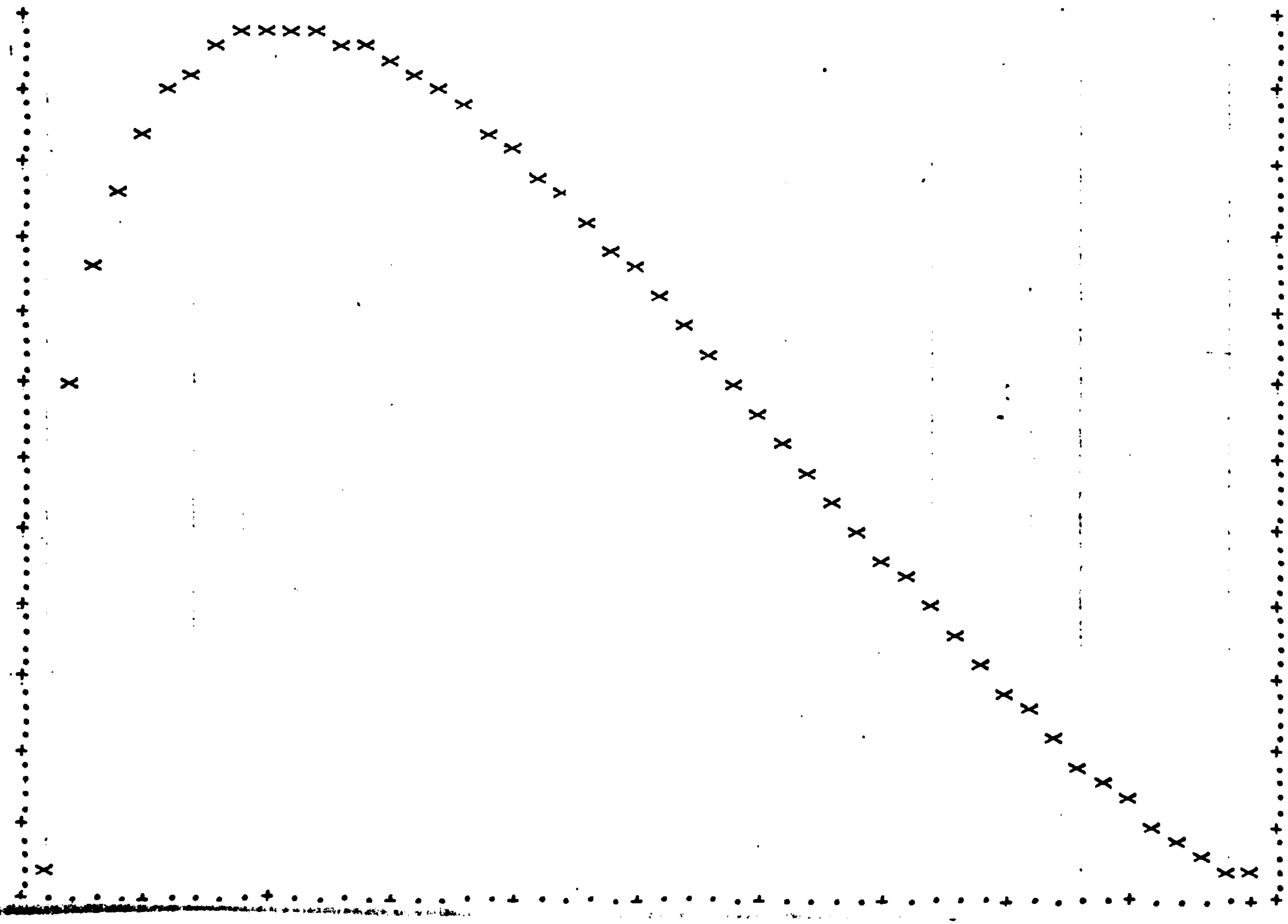
RIGHT-TO-LEFT
Normalized Mode = .85



LEFT-TO-RIGHT
Normalized Mode = .20

HIGH VARIANCE
Uncertainty Coefficient = .75

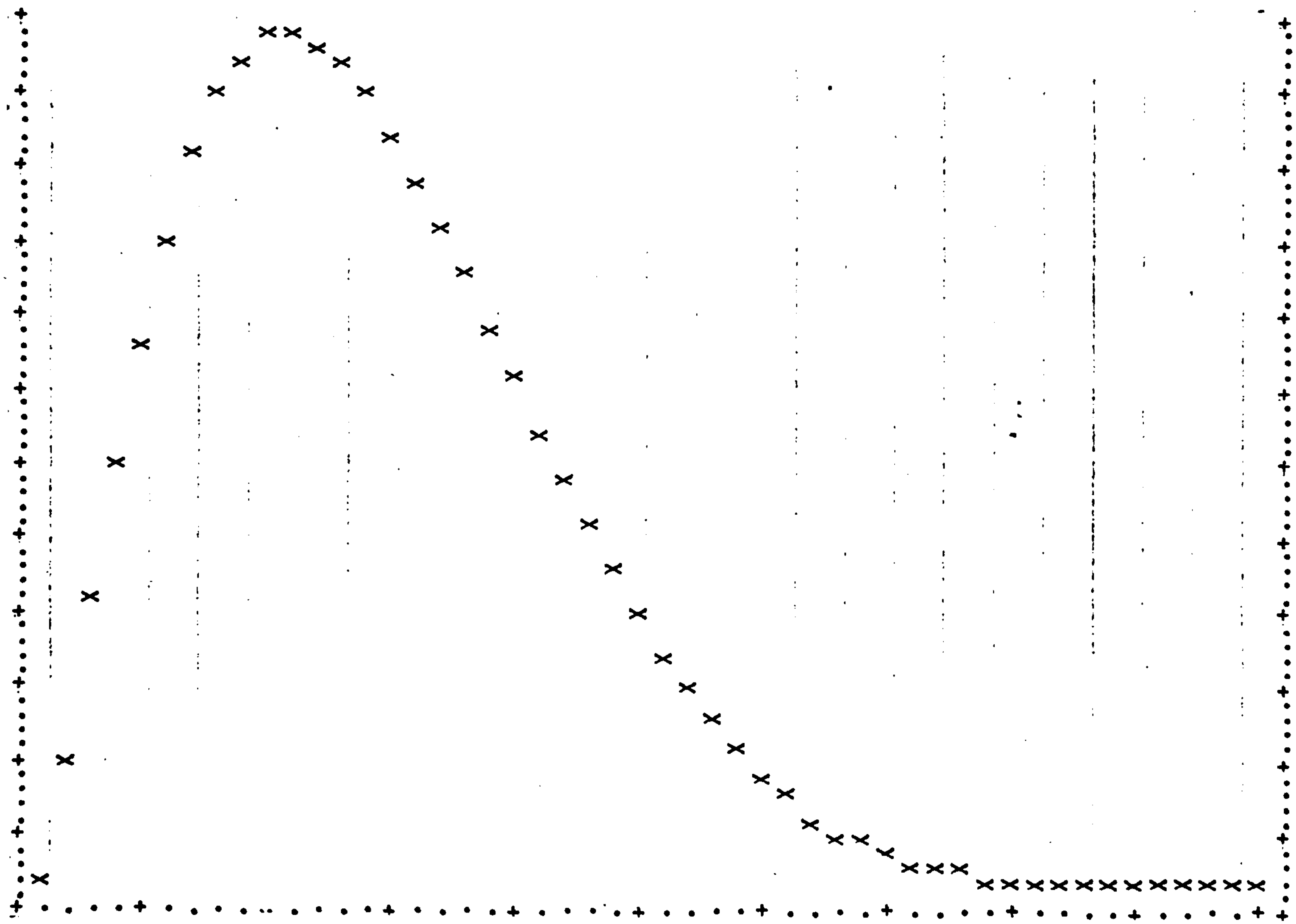
RIGHT-TO-LEFT
Normalized Mode = .80



LEFT-TO-RIGHT
Normalized Mode = .20

MEDIUM VARIANCE
Uncertainty Coefficient = .50

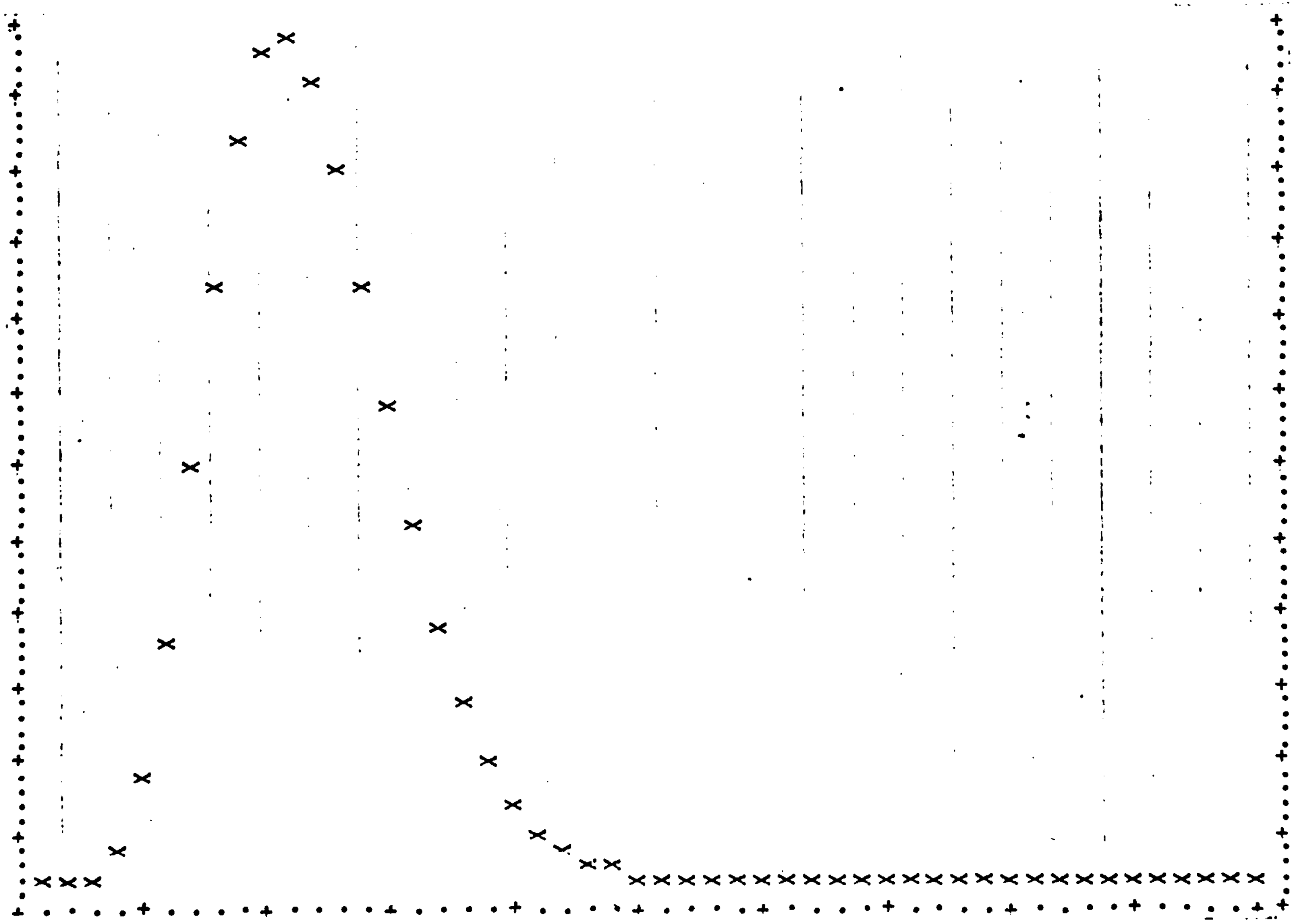
RIGHT-TO-LEFT
Normalized Mode = .80



LEFT-TO-RIGHT
Normalized Mode = .20

LOW VARIANCE
Uncertainty Coefficient = .25

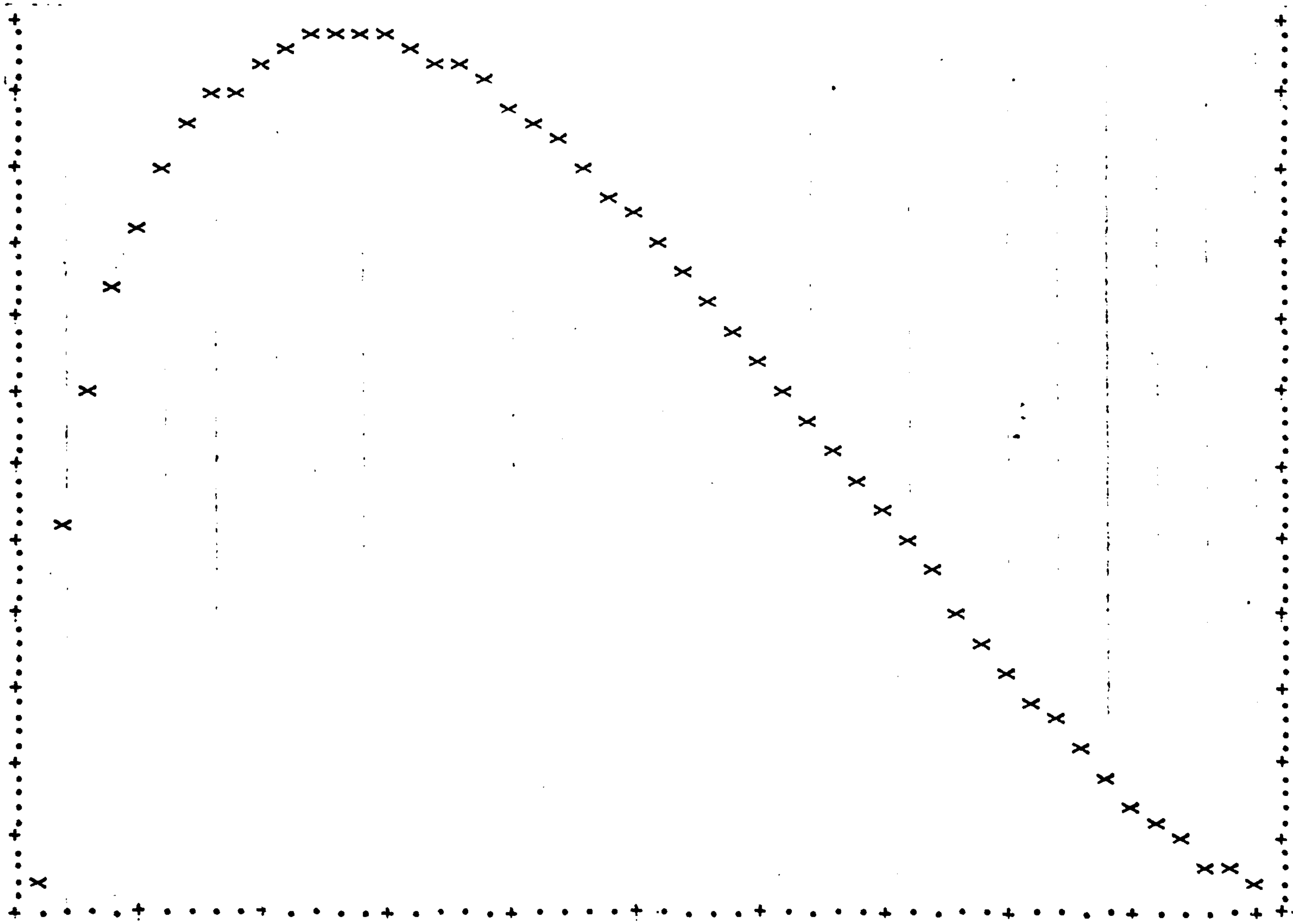
RIGHT-TO-LEFT
Normalized Mode = .80



LEFT-TO-RIGHT
Normalized Mode = .25

HIGH VARIANCE
Uncertainty Coefficient = .75

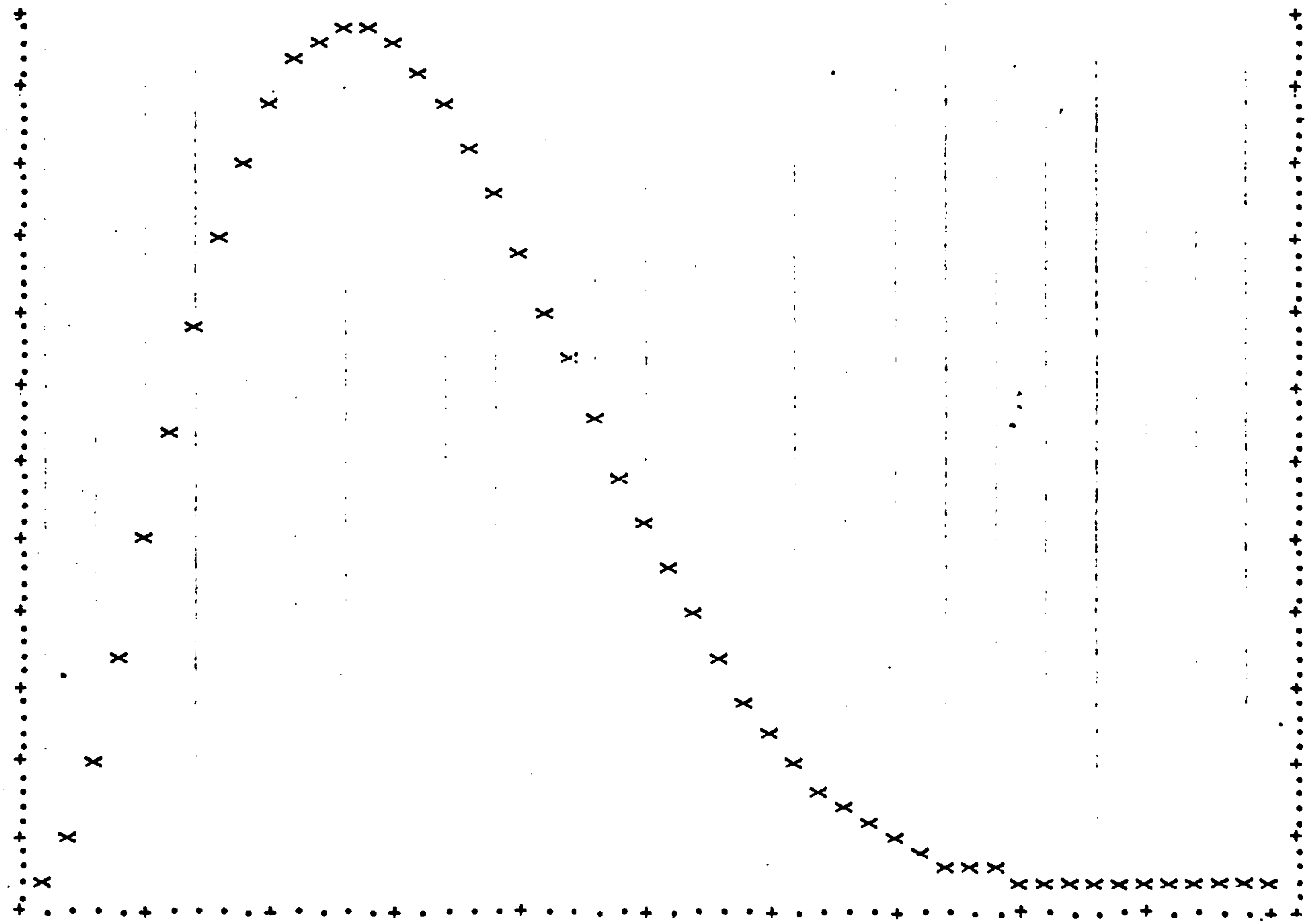
RIGHT-TO-LEFT
Normalized Mode = .75



LEFT-TO-RIGHT
Normalized MOde = .25

MEDIUM VARIANCE
Uncertainty Coefficient = .50

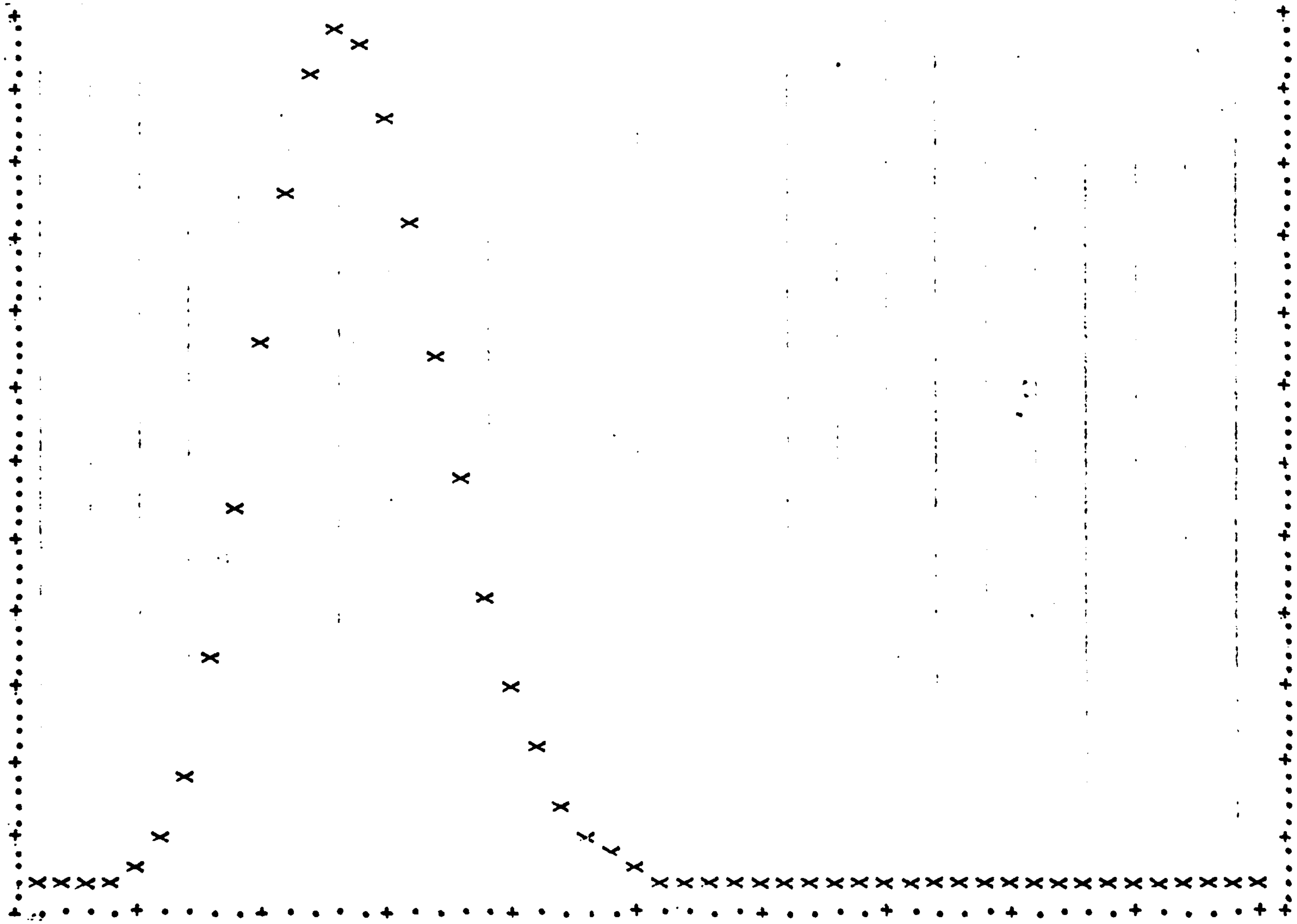
RIGHT-TO-LEFT
Normalized Mode = .75



LEFT-TO-RIGHT
Normalized Mode = .25

LOW VARIANCE
Uncertainty Coefficient = .25

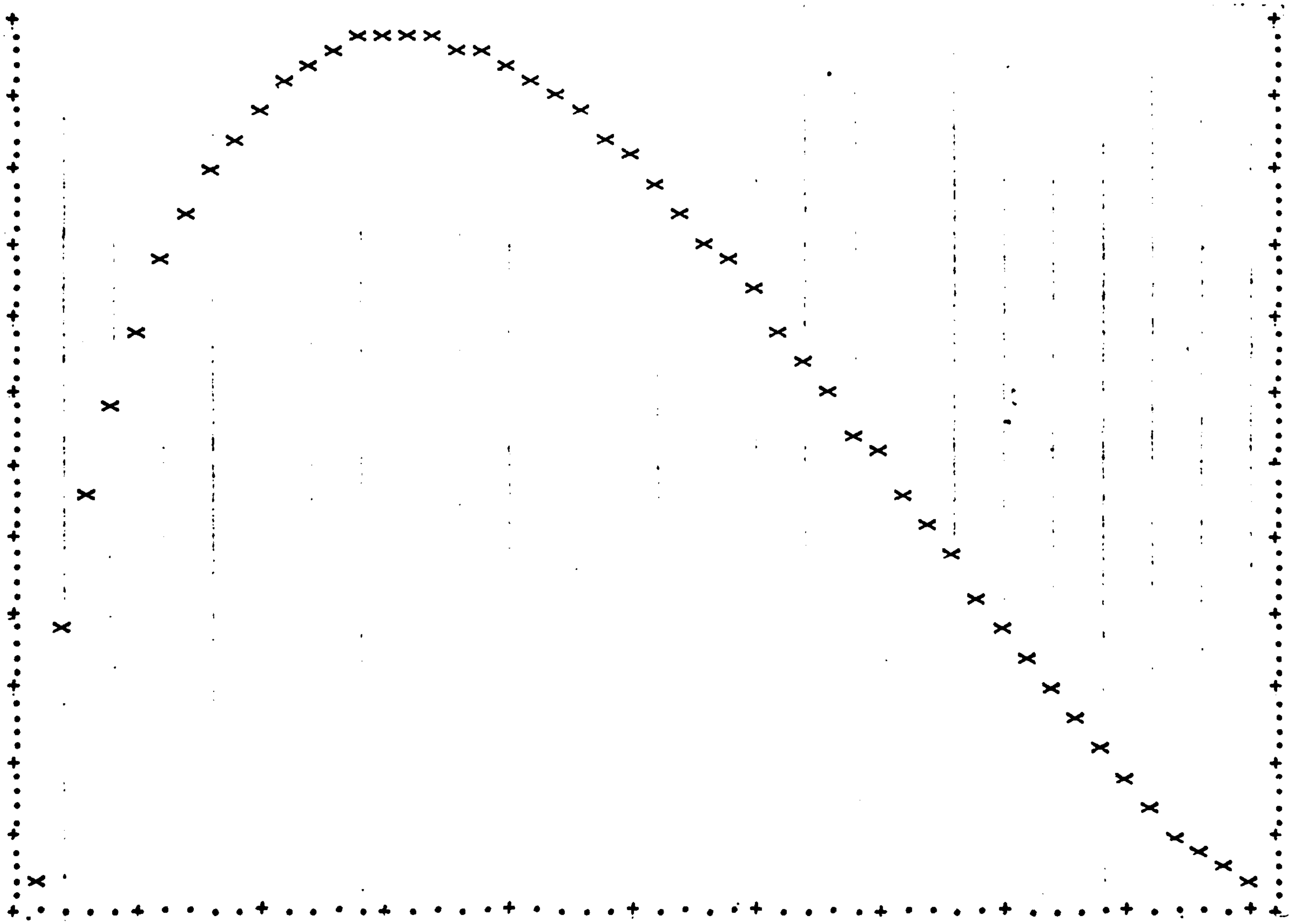
RIGHT-TO-LEFT
Normalized Mode = .75



LEFT-TO-RIGHT
Normalized Mode = .30

HIGH VARIANCE
Uncertainty Coefficient = .75

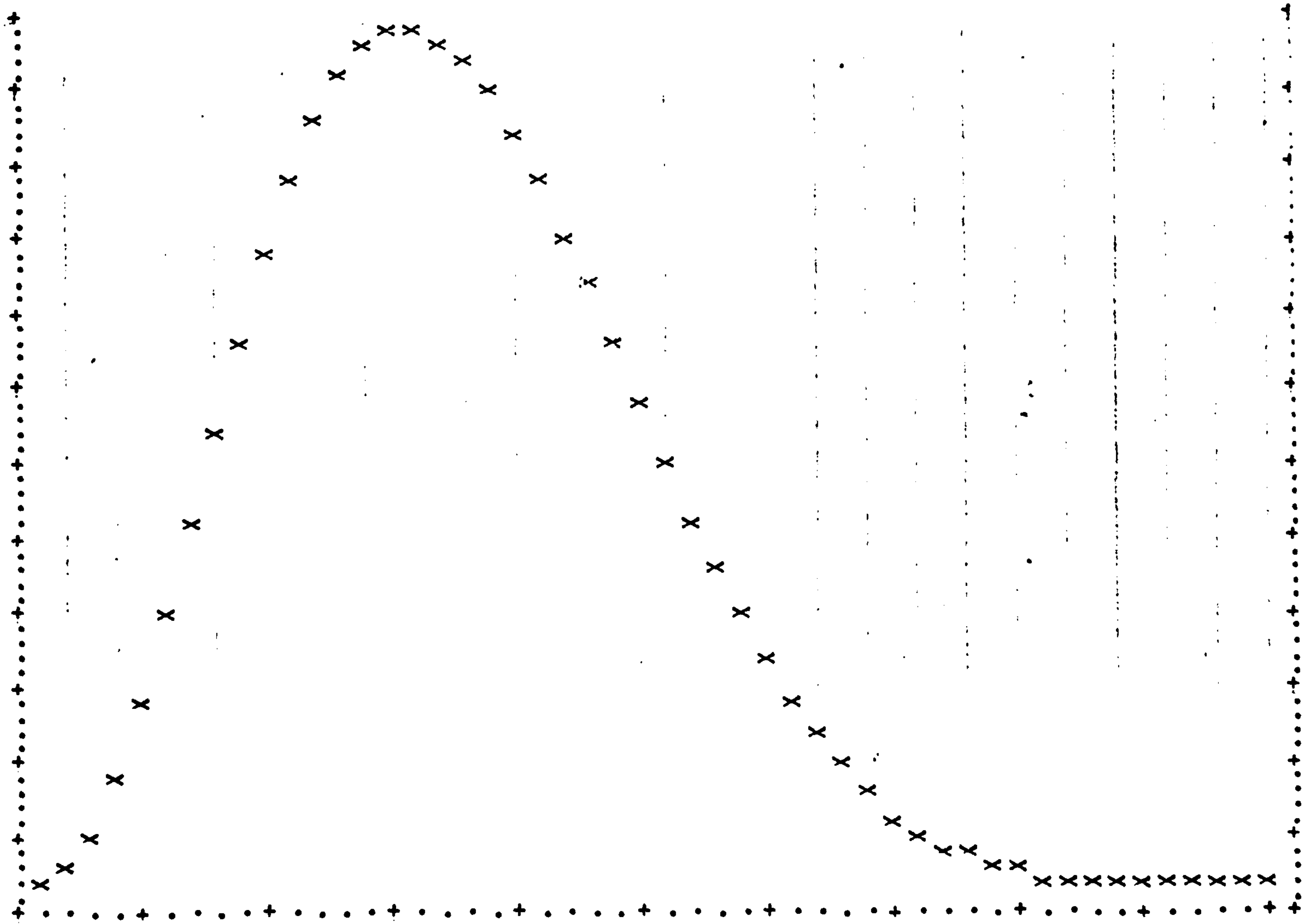
RIGHT-TO-LEFT
Normalized Mode = .70



LEFT-TO-RIGHT
Normalized Mode = .30

MEDIUM VARIANCE
Uncertainty Coefficient = .50

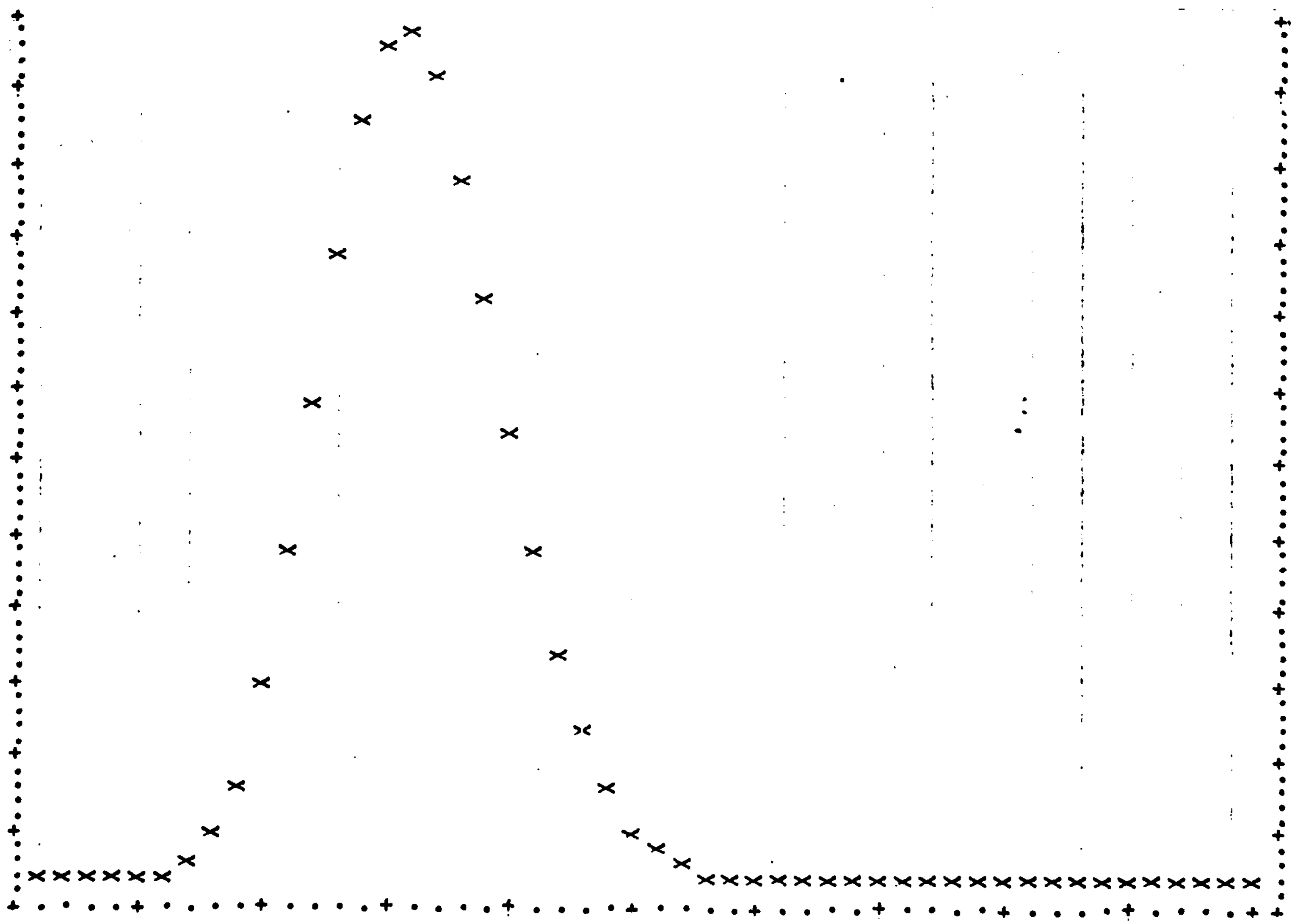
RIGHT-TO-LEFT
Normalized Mode = .70



LEFT-TO-RIGHT
Normalized Mode = .30

LOW VARIANCE
Uncertainty Coefficient = .25

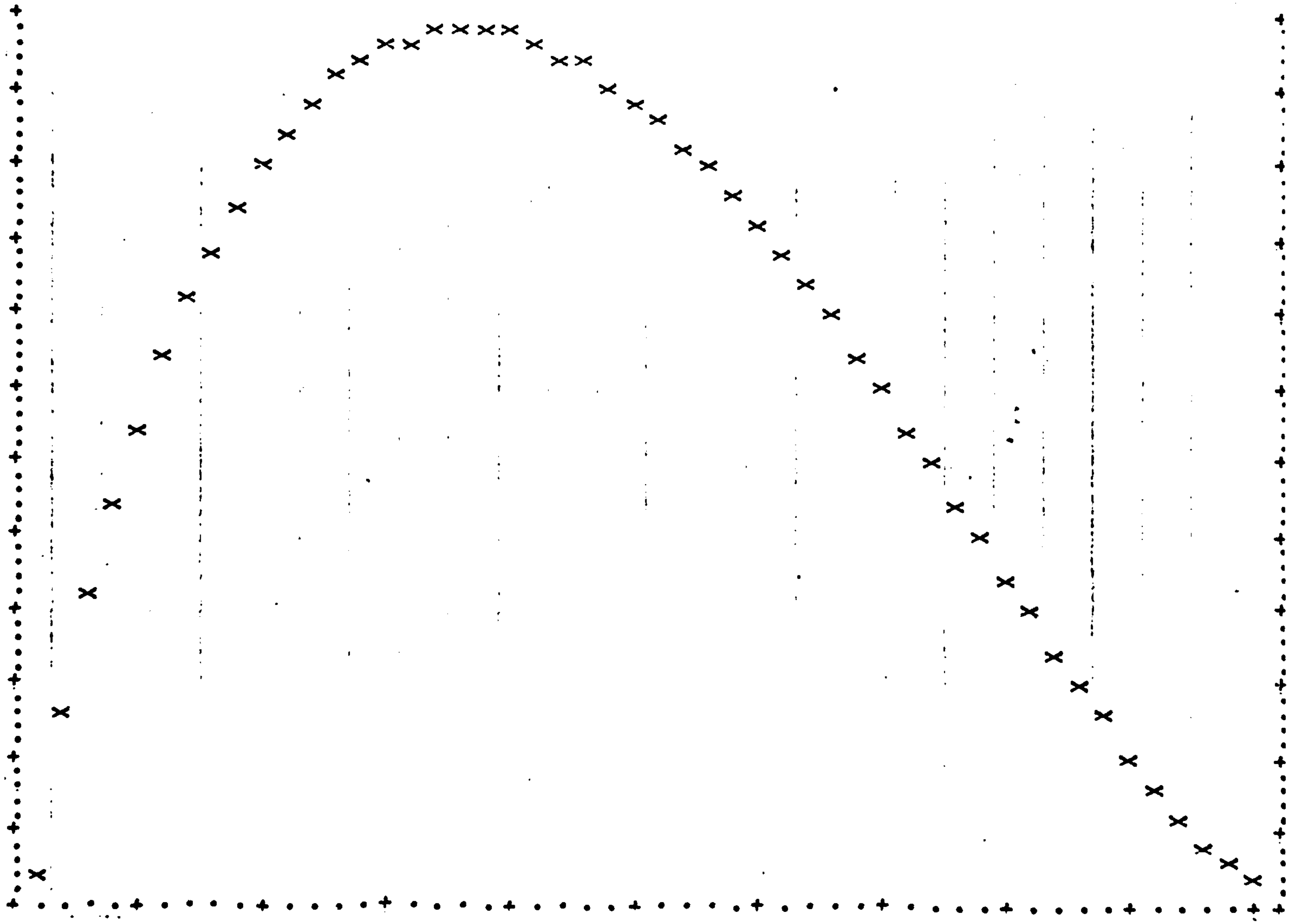
RIGHT-TO-LEFT
Normalized Mode = .70



<u>LEFT-TO-RIGHT</u>
Normalized Mode = .35

<u>HIGH VARIANCE</u>
Uncertainty Coefficient = .75

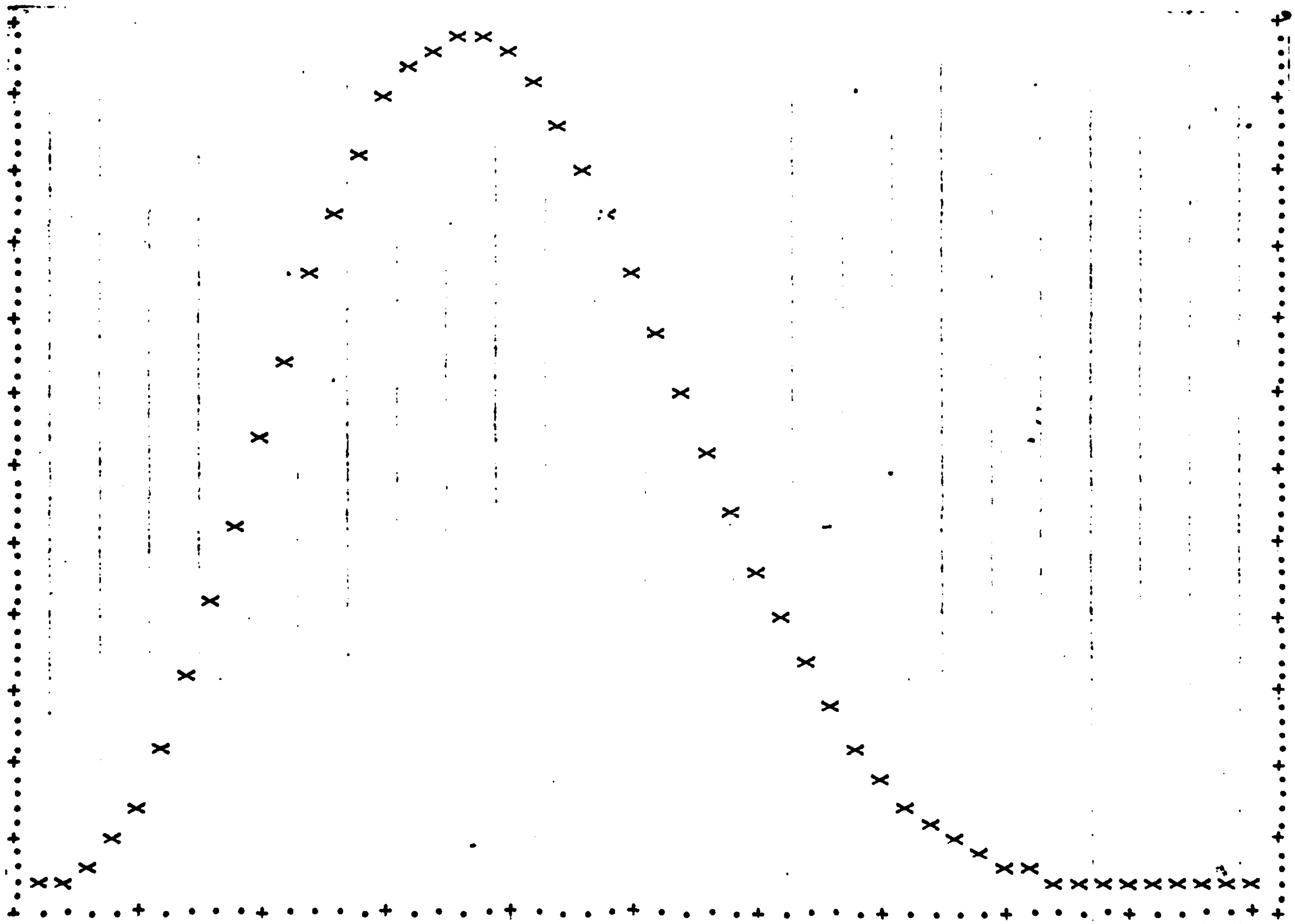
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .65



LEFT-TO-RIGHT
Normalized Mode = .35

MEDIUM VARIANCE
Uncertainty Coefficient = .50

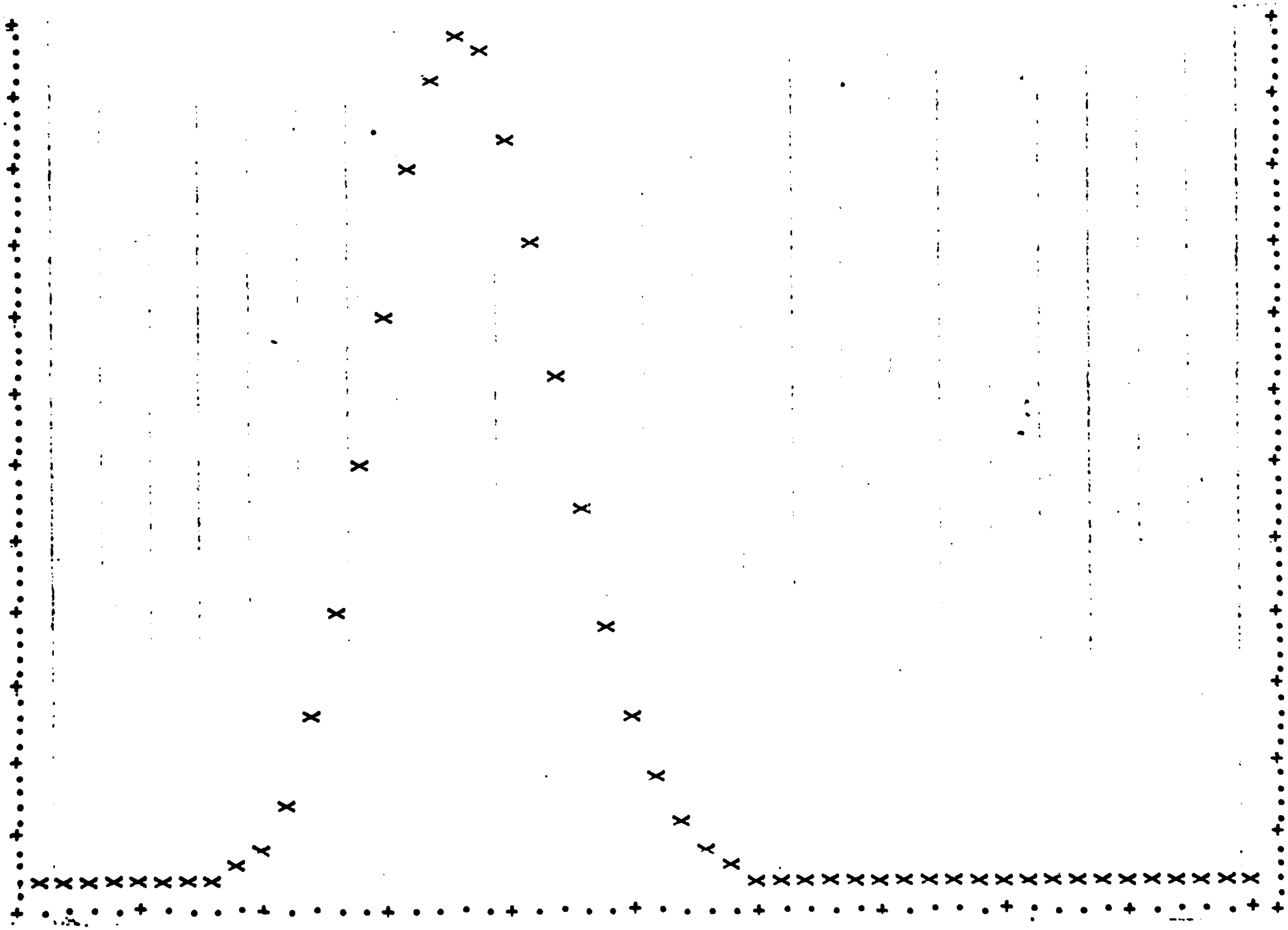
RIGHT-TO-LEFT
Normalized Mode = .65



LEFT-TO-RIGHT
Normalized Mode = .35

LOW VARIANCE
Uncertainty Coefficient = .25

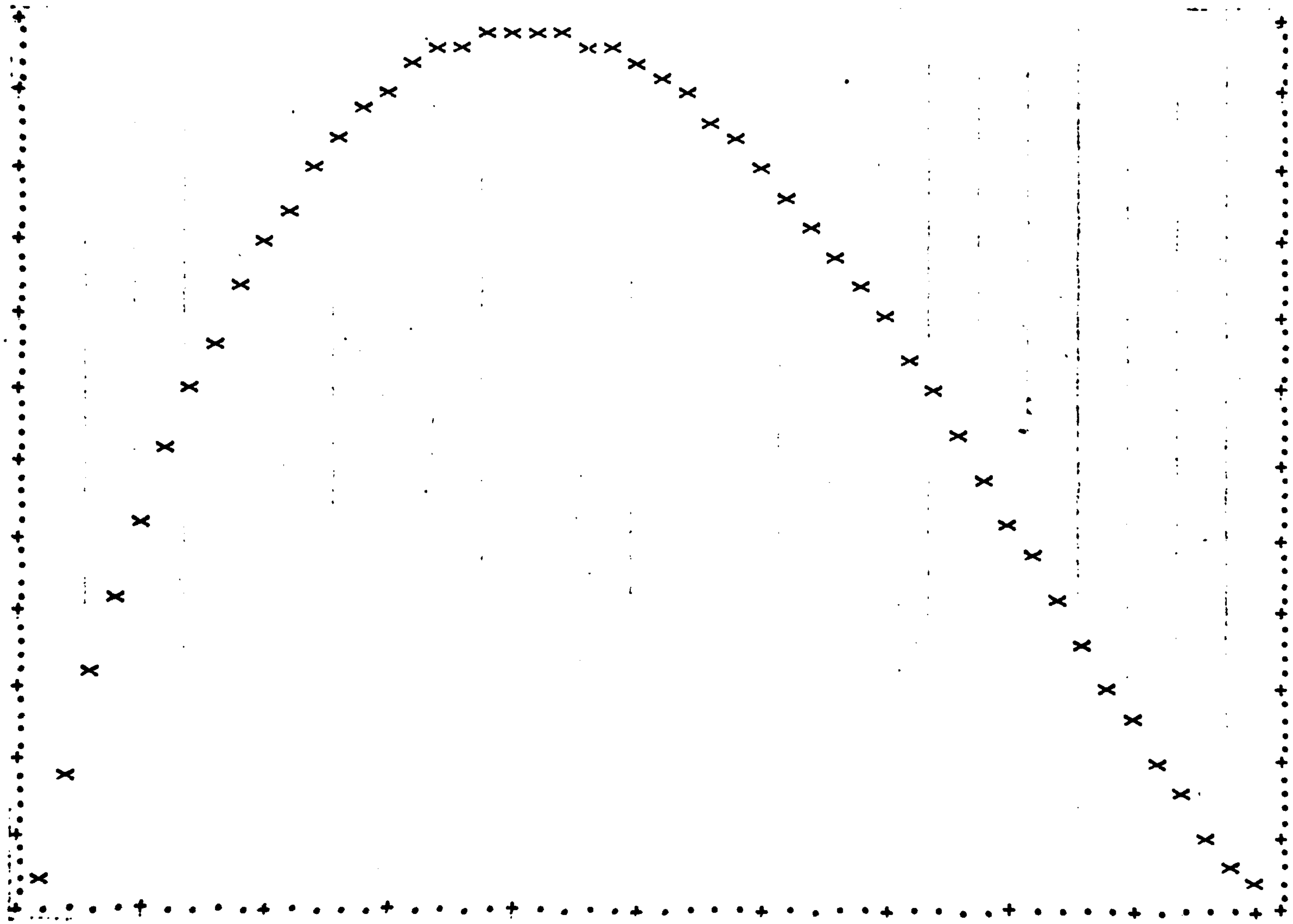
RIGHT-TO-LEFT
Normalized Mode = .65



LEFT-TO-RIGHT
Normalized Mode = .40

HIGH VARIANCE
Uncertainty Coefficient = .75

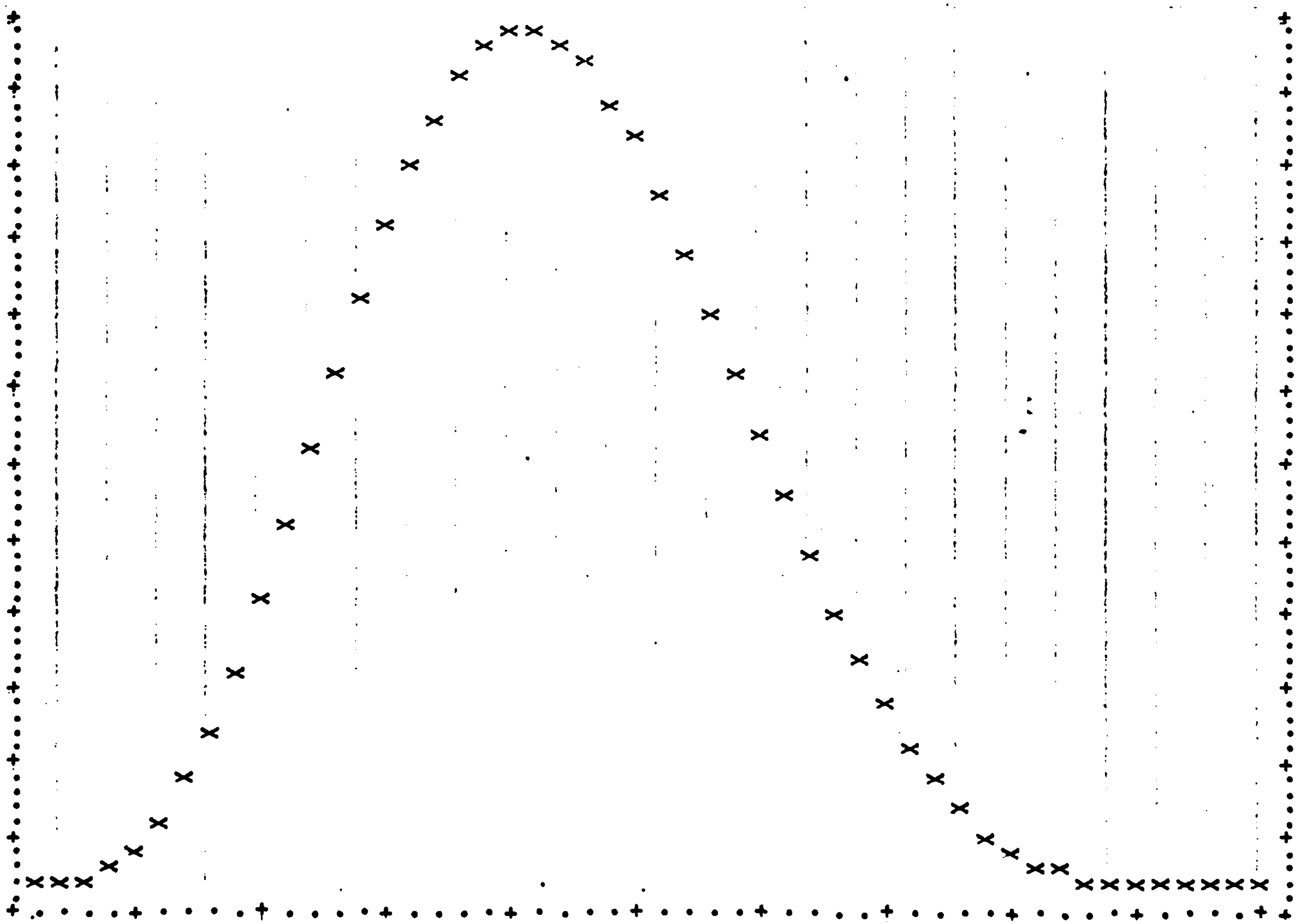
RIGHT-TO-LEFT
Normalized Mode = .60



LEFT-TO-RIGHT
Normalized Mode = .40

MEDIUM VARIANCE
Uncertainty Coefficient = .50

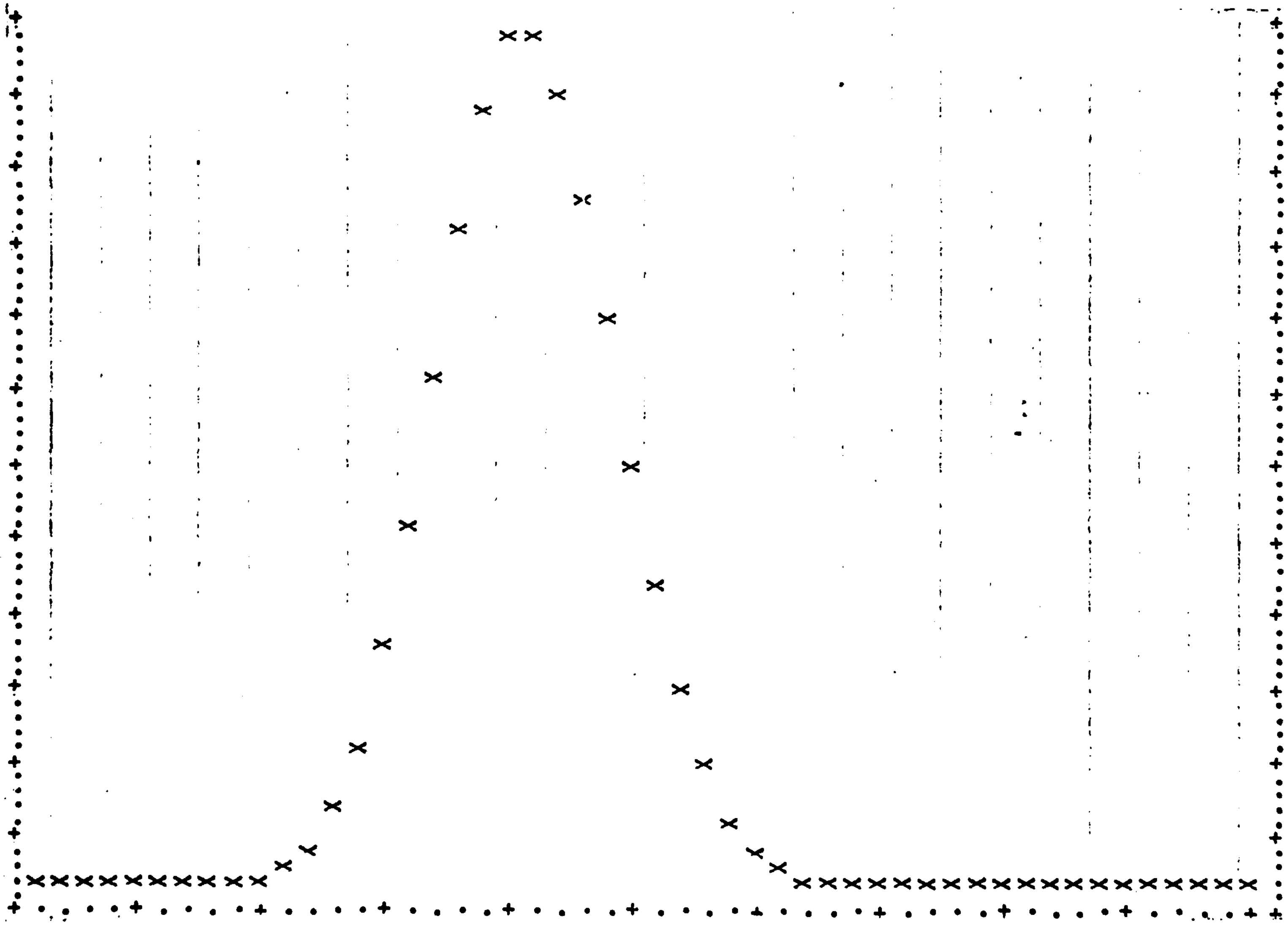
RIGHT-TO-LEFT
Normalized Mode = .60



<u>LEFT-TO-RIGHT</u>
Normalized Mode = .40

<u>LOW VARIANCE</u>
Uncertainty Coefficient = .25

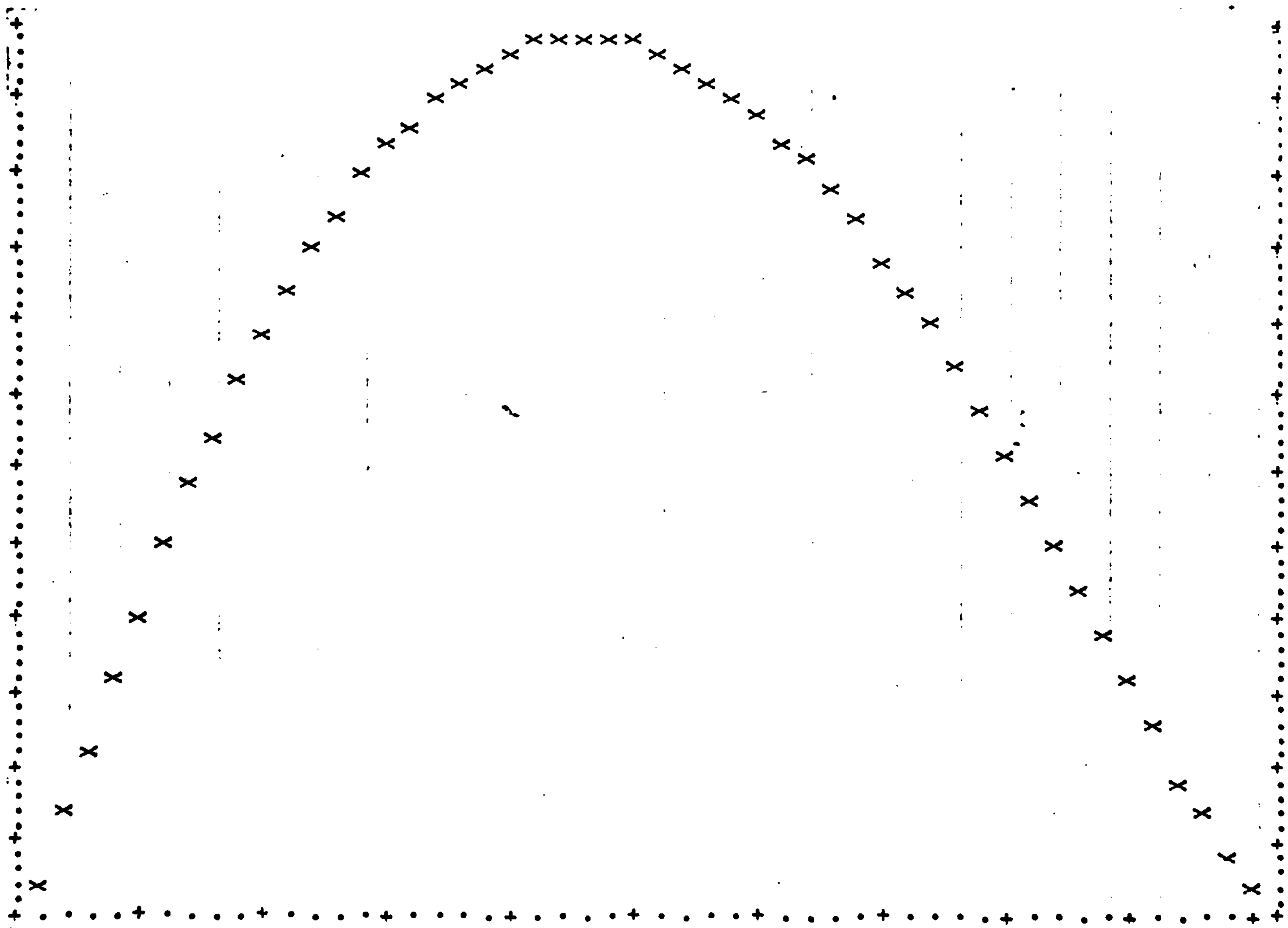
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .60



LEFT-TO-RIGHT
Normalized Mode = .45

HIGH VARIANCE
Uncertainty Coefficient = .75

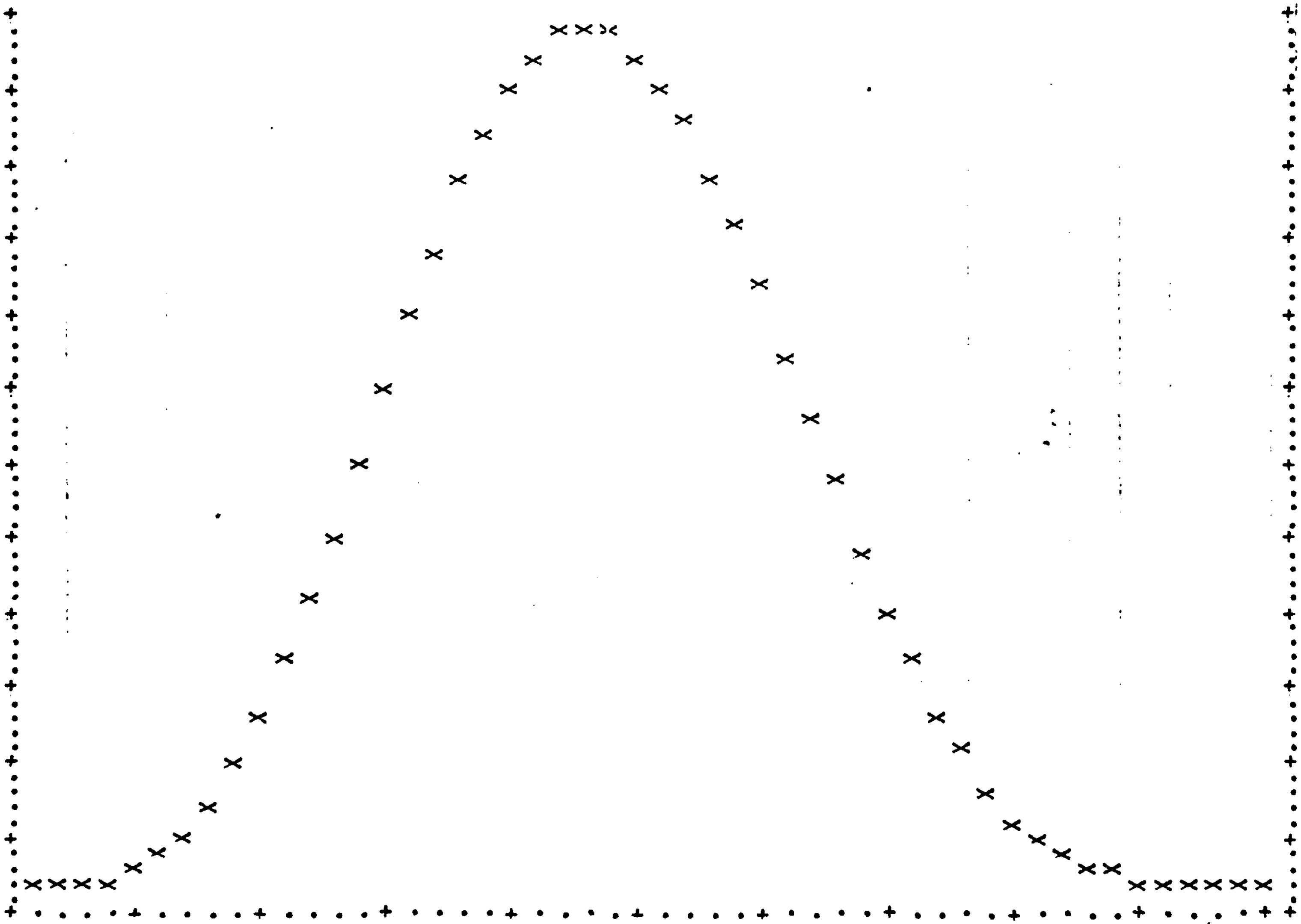
RIGHT-TO-LEFT
Normalized Mode = .55



LEFT-TO-RIGHT
Normalized Mode = .45

MEDIUM VARIANCE
Uncertainty Coefficient = .50

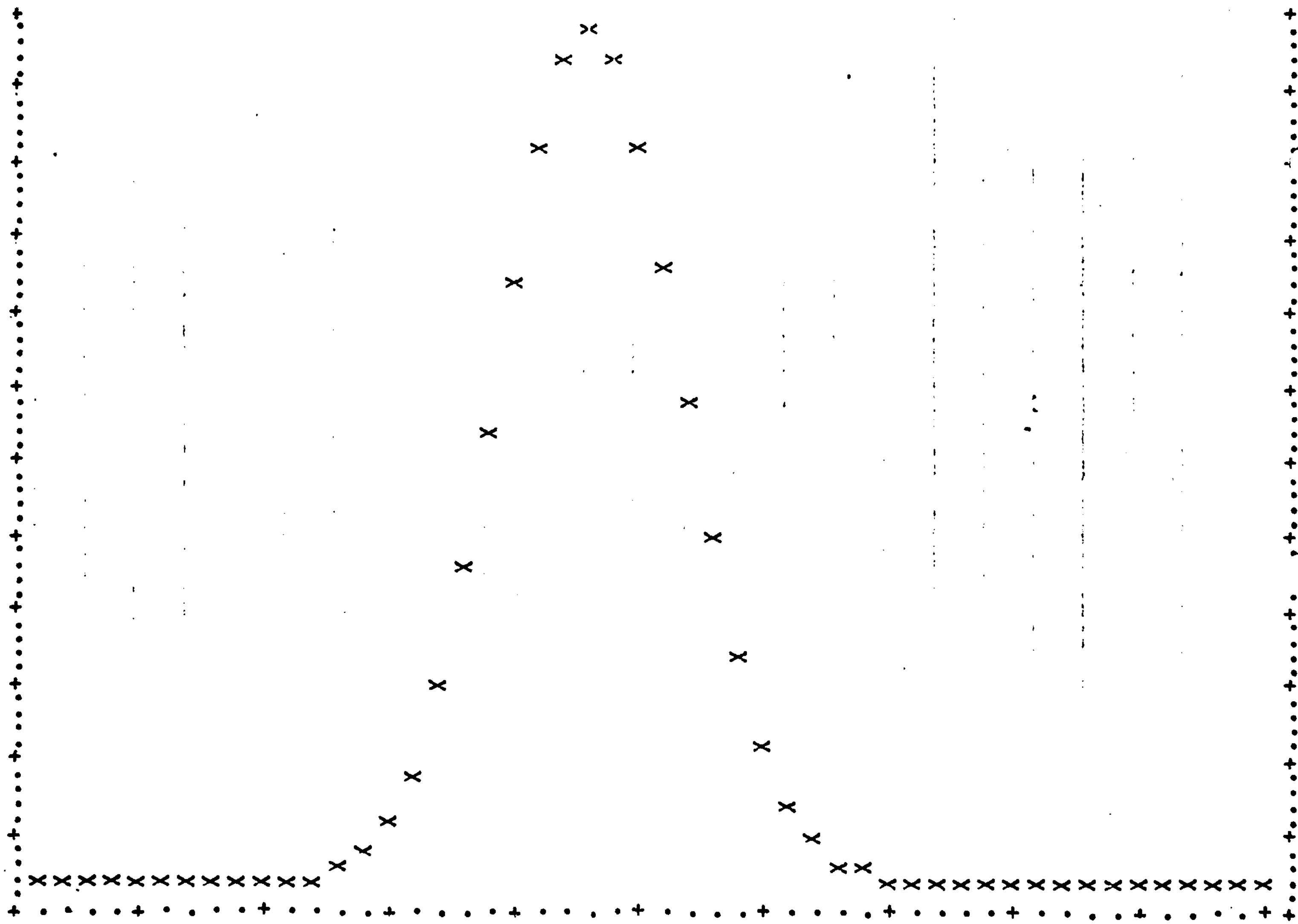
RIGHT-TO-LEFT
Normalized Mode = .55



LEFT-TO-RIGHT
Normalized Mode = .45

LOW VARIANCE
Uncertainty Coefficient = .25

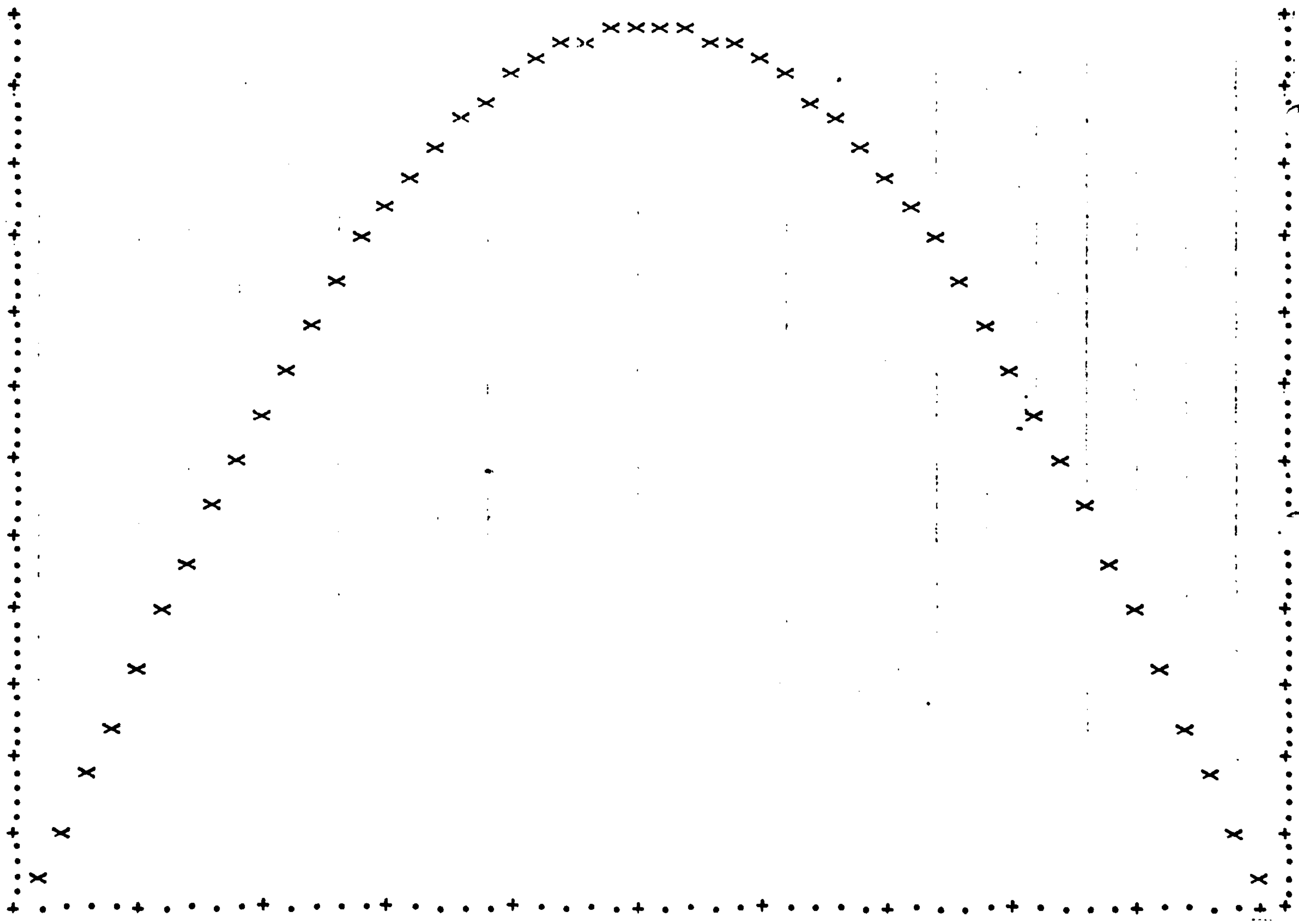
RIGHT-TO-LEFT
Normalized Mode = .55



<u>LEFT-TO-RIGHT</u>
Normalized Mode = .50

<u>HIGH VARIANCE</u>
Uncertainty Coefficient = .75

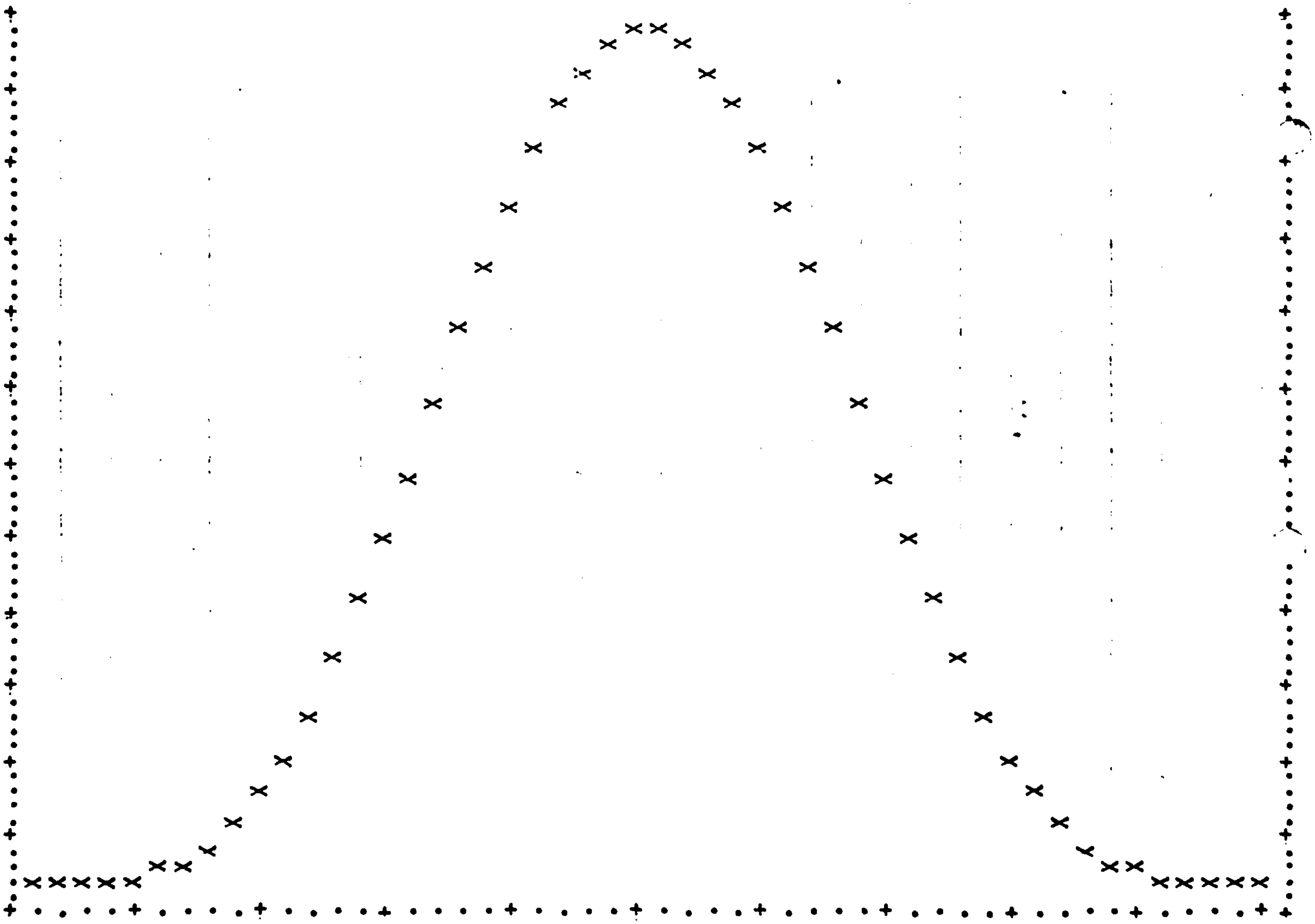
<u>RIGHT-TO-LEFT</u>
Normalized Mode = .50



<u>LEFT-TO-RIGHT</u>
<u>Normalized Mode</u> = .50

<u>MEDIUM VARIANCE</u>
<u>Uncertainty Coefficient</u> = .50

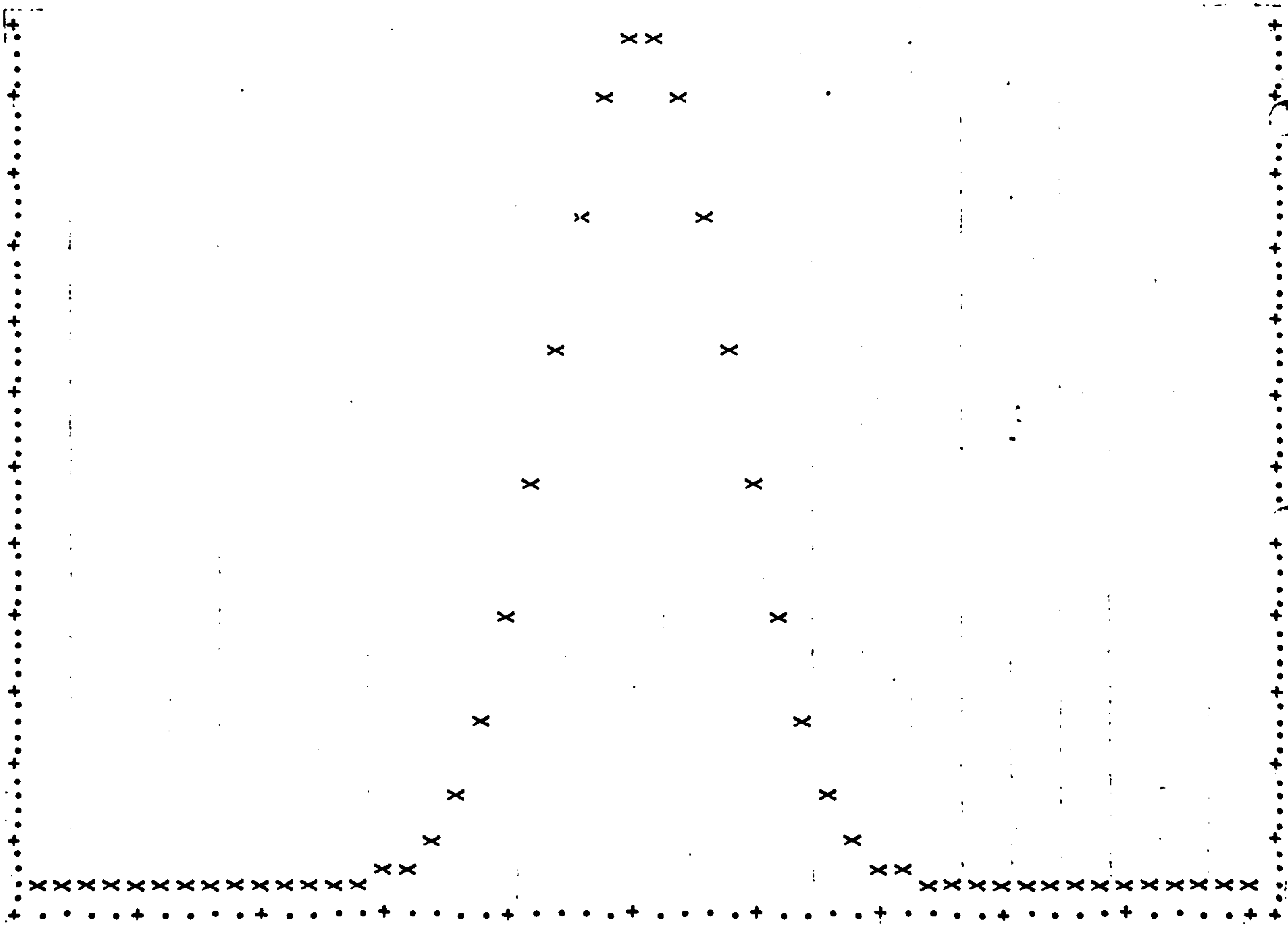
<u>RIGHT-TO-LEFT</u>
<u>Normalized Mode</u> = .50



LEFT-TO-RIGHT
Normalized Mode = .50

LOW VARIANCE
Uncertainty Coefficient = .25

RIGHT-TO-LEFT
Normalized Mode = .50



APPENDIX C

COMPUTER PROGRAM SPET

Description

Program SPET is written in FORTRAN IV for use in conjunction with a Univac 1108 computer in an interactive time-sharing mode. It is designed to approximate the frequency distribution of the sum of up to fifty independent beta or other random variables by means of Monte Carlo simulation. This is accomplished in the following manner:

(1) The user specifies the beta variables by entering the lowest (L), most likely (M), and highest (H) value as well as the uncertainty coefficient (U) of each variable. He specifies other variables by entering the lowest, most likely, and highest values, as well as the mean, variance, and cumulative density function (cdf) of each variable. He enters only those values of the cdf F(x) corresponding to x = L + i(H - L)/10, i = 1,...,10.

(2) For all beta variables the computer:

(a) Computes beta parameters a and b from M and U by first converting U to V with equation [6] and then simultaneously solving equations [3] and [4].

(b) Computes the discrete cdf, F(x), for x = i/10, i = 1,...,10, using Subroutine DQG32.1/

(3) Next the class intervals for the distribution of the sum of the beta and other variables are computed. Adding the L values and throughput (TPUT)2/ establishes the lower limit of the range; adding the H values and TPUT establishes the upper limit. The range is then divided into 15 intervals of equal width.3/

(4) Four frequency distributions of the sum are generated. The first distribution results from the assumption that the distributions making up the sum are statis-

---

1/ *DQG-32 uses the 32 point Gaussian quadrature method of integration. It is taken from* Convolution of Inverse Beta Distributions by a Sampling Technique *(Bethesda, MD: Mathematica, Inc. 1971).*

2/ *Throughput means constant, and is usually the cost of a subsystem that is known with certainty.*

3/ *The number of intervals can be varied by assigning the desired number to KK in line 11 of the main program.*

tically independent. It is generated as follows:

(a) Obtain a random number lying between zero and one from a uniform random number generator.4/

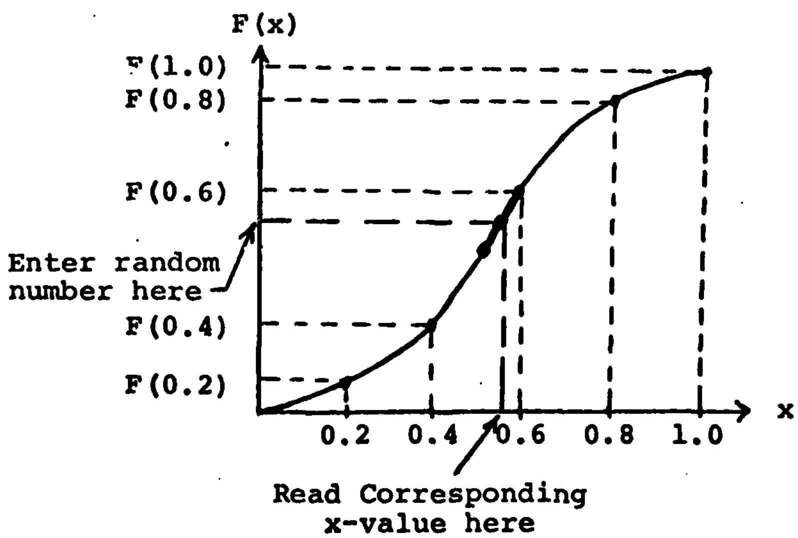
(b) Compare the generated number with the values of the discrete cdf, $F(x)$, for one of the variables and note the interval $[F(x_i),F(x_j)]$, $x_i < x_j$, into which it falls.

(c) Find the x-value, $x_i < x < x_j$, corresponding to the random number by means of linear interpolation.5/

(d) Transform the x-value from its normalized (0,1) value to its standard (L,H) value by means of the transformation $x^* = L + (H - L)x$. [C1]

4/ *An unsuccessful attempt was made to find a machine-independent random number generator for inclusion in the program. Therefore, Program SPET requires the use of a user-supplied generator. Make the appropriate changes in line 150 of the main program to accommodate the generator (it may be necessary to also change lines 68-69, 155, 158, 162, 175, 178, 182, 262, and 370-372).*

5/ *Steps (b) and (c) can be illustrated for a normalized random variable x as follows:*



(e) Repeat steps (a) - (d) for every variable.

(f) Compute the observation,

$$X_j = TPUT + \sum_{i=1}^{N} x_i^*$$

where N denotes the number of random variables.

(g) Find the class interval [see (3) above] in which X_j occurs and register one occurrence in that interval.

(h) Repeat steps (a) - (g) for j = 1,...,M observations, where M = number of desired observations. When step (h) is terminated, one has a frequency distribution of the sum of independent random variables.

(5) The second distribution of the sum is generated in the same way as the first except for step (e), which should now read, "Repeat steps (b) - (d) for every variable." This means that the same random number is used to obtain an observation on each component of the sum rather than a new number as done when constructing the first distribution. The procedure of using only one random number introduces a correlation among the component variables because when the value of one variable is known, it in turn maps uniquely to the values of all other variables. This correlation is positive because the cdfs are positive monotonic functions.

(6) The third distribution is simply the uniform distribution over the range of the first distribution. Both the second and third distributions serve as indicators of how a violation of the independence assumption could affect the distribution of the sum and the summary statistics.

(7) The fourth distribution is the same as the first distribution except the component variables are all uniform random variables. This distribution serves as an indicator of the relative sensitivity of the distribution of the sum to the degree of uncertainty in the component variables.

(8) Along with the four frequency distributions just described, the computer generates the mean, mode, standard deviation, variance, 90% confidence interval about the

mean, and the probability of exceeding the mean for each of the distributions. In addition, the user can specify any confidence interval about the mean and the probability of exceeding any number within the range of the distribution and the computer will generate it for him.

User Instructions

The following example illustrates the use of Program SPET:

```
!SPET

TITLE
**EXAMPLE**
NUMBER OF BETA DISTRIBUTIONS
4
DATA FILE(1) OR TERMINAL(2) INPUT
2
LOWEST,MODE,HIGHEST,UNCERTAINTY COEFF
DATA
100,250,900,.8
DATA
300,350,390,.2
DATA
200,400,675,.5
DATA
150,330,800,.4
NUMBER OF OTHER DISTRIBUTIONS
1
LOWEST,MODE,HIGHEST,MEAN,VARIANCE,DISCRETE CDF
DATA
50,75,100,75,208.3,.1,.2,.3,.4,.5,.6,.7,.8,.9,1.0
THROUGHPUT
35
NUMBER OF OBSERVATIONS
2000
SEED RANDOM NUMBER GENERATOR
87654321
```

EXAMPLE

I N P U T S

OBSERVATIONS = 2000
SEED = 87654321

BETA VBLE	LOWEST	MODE	HIGHEST	U COEFF	NMODE	ALPHA	BETA	MEAN	VARIANCE
1	100.0	250.0	900.0	.80	.19	.3	1.1	396.4	33806.0
2	300.0	350.0	390.0	.20	.56	39.6	31.6	349.9	27.0
3	200.0	400.0	675.0	.50	.42	3.7	5.1	406.9	4697.4
4	150.0	330.0	800.0	.40	.28	3.6	9.4	349.3	5614.6
OTHER VBLE									
5	50.0	75.0	100.0					75.0	208.3
THROUGHPUT	35.0		35.0					35.0	
SUMS	835.0		2900.0					1612.5	44353.2

OTHER VBLE	CDF									
5	.1000	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9000	1.0000

OUTPUT

INTERVAL	RANGE		INDEPENDENT BETA/OTHER PDF	INDEPENDENT BETA/OTHER CDF	DEPENDENT BETA/OTHER PDF	DEPENDENT BETA/OTHER CDF	TOTAL UNIFORM PDF	TOTAL UNIFORM CDF	INDEPENDENT UNIFORM PDF	INDEPENDENT UNIFORM CDF
1	835.0 -	972.7	0.	0.	.0110	.0110	.0667	.0667	.0010	.0010
2	972.7 -	1110.3	.0010	.0010	.0545	.0655	.0667	.1333	.0025	.0035
3	1110.3 -	1248.0	.0240	.0250	.1040	.1695	.0667	.2000	.0215	.0250
4	1248.0 -	1385.7	.1140	.1390	.1245	.2940	.0667	.2667	.0435	.0685
5	1385.7 -	1523.3	.2270	.3660	.1355	.4295	.0667	.3333	.0805	.1490
6	1523.3 -	1661.0	.2360	.6020	.1365	.5660	.0667	.4000	.1140	.2630
7	1661.0 -	1798.7	.1945	.7965	.1170	.6830	.0667	.4667	.1575	.4205
8	1798.7 -	1936.3	.1240	.9205	.1115	.7945	.0667	.5333	.1615	.5820
9	1936.3 -	2074.0	.0600	.9805	.0875	.8820	.0667	.6000	.1525	.7345
10	2074.0 -	2211.7	.0185	.9990	.0540	.9360	.0667	.6667	.1070	.8415
11	2211.7 -	2349.3	.0010	1.0000	.0385	.9745	.0667	.7333	.0895	.9310
12	2349.3 -	2487.0	0.	1.0000	.0175	.9920	.0667	.8000	.0415	.9725
13	2487.0 -	2624.7	0.	1.0000	.0065	.9985	.0667	.8667	.0210	.9935
14	2624.7 -	2762.3	0.	1.0000	.0015	1.0000	.0667	.9333	.0060	.9995
15	2762.3 -	2900.0	0.	1.0000	0.	1.0000	.0667	1.0000	.0005	1.0000

	INDEPENDENT BETA/OTHER	DEPENDENT BETA/OTHER	TOTAL UNIFORM	INDEPENDENT UNIFORM
MEAN	1615.5	1621.1	1867.5	1869.2
MODE	1592.2	1592.2		1867.5
VARIANCE	44843.7	128488.0	355352.1	103015.0
STD DEVIATION	211.8	358.5	596.1	321.0
90% CONFIDENCE INTERVAL	1307.2, 2059.2	1124.8, 2364.7	938.3, 2796.8	1337.5, 2423.2
PROB EXCEED MEAN	.48	.47	.50	.50

ANOTHER CONFIDENCE INTERVAL?
95

95% CONFIDENCE INTERVAL	1277.0, 2218.5	1071.2, 2900.0	886.6, 2848.4	1258.3, 2524.7

ANOTHER CONFIDENCE INTERVAL?
0

PROB EXCEED SOME VALUE?
1450

PROB EXCEED 1450.0	.75	.64	.70	.89

PROB EXCEED SOME VALUE?
0

ADDITIONAL OBSERVATIONS?
0

ANOTHER SEED?
0
STOP
SRU'S114.7

(1) The user enters a title up to 60 characters long.

(2) If he is using beta distributions to represent some or all of his component distributions, the user enters the number of distributions to be so represented and proceeds to step (a) below. However, if he is not using the beta, he enters the number "0" and proceeds to step (3).

(a) The user specifies whether he will enter the four-tuples L, M, H, U defining each beta variable directly from the terminal or from a data file stored in the computer by entering the number "1" for data file input or the number "2" for terminal input.

(b) If he chooses the terminal input, his next step is to enter one four-tuple L, M, H, U for each beta variable. If he chooses the data file input, he merely enters the name of the data file.

(3) If he is using distributions other than the beta to represent some or all of his component distributions, the user enters the number of distributions to be so represented, followed by the L, M, H, mean, variance, and the discrete cdf (see page C-1) of the distributions he has chosen. If he is not using other distributions, he enters the number "0".

(4) If there is a throughput (constant), the user enters it now.

(5) He then enters the number of observations (sample size) he desires, followed by a seed for the random number generator.

(6) The computer prints the user's inputs, followed by the output.

(7) The computer then queries the user if he desires another confidence interval. The user responds with the confidence interval he desires, or types the number "0" if he desires none.

(8) The computer asks the user if he desires the probability that some value within the range of the distribution will be exceeded. The user responds with that number, or the number "0" if he desires none.

(9) The computer inquires to see if the user desires additional observations. The user responds with the num-

ber of additional observations he desires (the number "0" if none). If he desires additional observations, the computer repeats steps (6) - (9).

(10) The computer asks the user if he desires to use another seed for the random number generator. The user responds with the seed if he does, or with the number "0" if he does not. If he enters another seed, the computer repeats steps (6) - (10).

Program Listing

```
DOUBLE PRECISION A(50),B(50),XL,XU,Y,ALPHA,BETA
REAL L(50),H(50),MODE(50),NM(50),U(50),C(50),MEAN(50),VAR(50)
REAL TITLE(12),VALINT(10,50),W(51),P1(15),P2(15),P3(15),P4(15)
REAL C1(15),C2(15),C3(15),C4(15)
REAL MODE1,MEAN1,MODE2,MEAN2,MEAN3,MODE4,MEAN4,MSUM,L1,L2,L3,L4
REAL NAME(4)
INTEGER F1(15),F2(15),F4(15)
DATA NAME/'!EQU','ATE ','3   ','    '/
II=6
JJ=5
KK=15
%
% INPUT
%
WRITE(II,5)
5 FORMAT(//6H TITLE)
READ(JJ,7) TITLE
7 FORMAT(15A4)
WRITE(II,10)
10 FORMAT(1H ,28HNUMBER OF BETA DISTRIBUTIONS)
READ(JJ,*) N1
IF(N1.EQ.0) GO TO 60
WRITE(II,11)
11 FORMAT(1H ,32HDATA FILE(1) OR TERMINAL(2) INPUT)
READ(JJ,*) INPUT
IF(INPUT.EQ.2) GO TO 16
WRITE(II,13)
13 FORMAT(1H ,17HNAME OF DATA FILE)
READ(JJ,14) NAME(4)
14 FORMAT(A4)
CALL OBEY(NAME,4)
DO 15 I=1,N1
READ(3,*) L(I),MODE(I),H(I),U(I)
15 CONTINUE
GO TO 60
16 CONTINUE
WRITE(II,30)
30 FORMAT(1H ,37HLOWEST,MODE,HIGHEST,UNCERTAINTY COEFF)
DO 40 I=1,N1
WRITE(II,50)
50 FORMAT(1H 4HDATA)
READ(JJ,*) L(I),MODE(I),H(I),U(I)
40 CONTINUE
60 CONTINUE
WRITE(II,70)
70 FORMAT(1H ,29HNUMBER OF OTHER DISTRIBUTIONS)
READ(JJ,*) N2
IF(N2.EQ.0) GO TO 78
NONE=N1+1
NTWO=N1+N2
WRITE(II,75)
75 FORMAT(1H ,46HLOWEST,MODE,HIGHEST,MEAN,VARIANCE,DISCRETE CDF)
DO 76 I=NONE,NTWO
WRITE(II,50)
READ(JJ,*) L(I),MODE(I),H(I),MEAN(I),VAR(I),(VALINT(J,I),J=1,10)
76 CONTINUE
GO TO 79
78 CONTINUE
NTWO=N1
79 WRITE(II,72)
72 FORMAT(1H ,10HTHROUGHPUT)
READ(JJ,*) TPUT
WRITE(II,77)
77 FORMAT(1H ,22HNUMBER OF OBSERVATIONS)
READ(JJ,*) M
WRITE(II,80)
80 FORMAT(1H ,28HSEED RANDOM NUMBER GENERATOR)
READ(JJ,*) KSD
KSD2=KSD
IF(N1.EQ.0) GO TO 105
%
% COMPUTE A AND B
%
DO 85 I=1,N1
N=0
```

```
NM(I)=(MODE(I)-L(I))/(H(I)-L(I))
V=(.28867508*U(I))**2.
B(I)=1.0
86 A(I)=B(I)*NM(I)/(1.-NM(I))
VBETA=((A(I)+1.)*(B(I)+1.))/(((A(I)+B(I)+2.)**2.)*(A(I)+B(I)+3.))
Z=V-VBETA
IF(N.EQ.1) GO TO 87
IF(Z) 3,85,1
87 IF(Z) 4,85,85
3 B(I)=B(I)+1.
GO TO 86
1 B(I)=B(I)-1.
N=1
GO TO 86
4 B(I)=B(I)+.05
GO TO 86
85 CONTINUE
%
% COMPUTE DISCRETE CDFS
%
XL=0.D0
DO 90 J=1,N1
XU=1.D0
ALPHA=A(J)
BETA=B(J)
CALL DQG32(XL,XU,Y,ALPHA,BETA)
C(J)=1./Y
XU=0.D0
VALINT(10,J)=1.
DO 100 I=1,9
XU=XU+.1D0
CALL DQG32(XL,XU,Y,ALPHA,BETA)
VALINT(I,J)=C(J)*Y
100 CONTINUE
90 CONTINUE
%
% COMPUTE CLASS INTERVALS
%
105 CONTINUE
XLOW=TPUT
XHIGH=TPUT
DO 110 I=1,NTWO
XLOW=XLOW+L(I)
XHIGH=XHIGH+H(I)
110 CONTINUE
RANGE=XHIGH-XLOW
WIDTH=RANGE/KK
W(1)=XLOW
W(KK+1)=XHIGH
DO 120 I=2,KK
W(I)=W(I-1)+WIDTH
120 CONTINUE
%
% GENERATE HISTOGRAM
%
140 CONTINUE
TOTAL1=0.
TOTAL2=0.
TOTAL4=0.
SUM1=0.
SUM2=0.
SUM4=0.
SUMSQ1=0.
SUMSQ2=0.
SUMSQ4=0.
MSTAR=M
DO 150 I=1,KK
F1(I)=0
F2(I)=0
F4(I)=0
150 CONTINUE
160 CONTINUE
DO 180 K=1,MSTAR
DO 190 J=1,NTWO
RAND=UDRNRT(KSD)
```

```
      IF(J.GT.1)GO TO 195
      DO 200 LL=1,NTWO
      DO 205 I=1,10
      MM=I
      IF(RAND.LE.VALINT(I,LL)) GO TO 210
  205 CONTINUE
  210 IF(MM.EQ.1) GO TO 215
      T2=(RAND-VALINT(MM-1,LL))/(VALINT(MM,LL)-VALINT(MM-1,LL))
      S=(MM-1)/10.
      SMALL2=.1*T2+S
      GO TO 220
  215 SMALL2=.1*(RAND/VALINT(1,LL))
  220 TOTAL2=TOTAL2+SMALL2*(H(LL)-L(LL))+L(LL)
  200 CONTINUE
      TOTAL2=TOTAL2+TPUT
      SUM2=SUM2+TOTAL2
      SUMSQ2=SUMSQ2+TOTAL2**2.
      COST2=(TOTAL2-XLOW)/RANGE
      J2=COST2*KK+1.
      IF(J2.EQ.KK+1.) J2=KK
      F2(J2)=F2(J2)+1.
      TOTAL2=0.
  195 DO 225 I=1,10
      MM=I
      IF(RAND.LE.VALINT(I,J)) GO TO 227
  225 CONTINUE
  227 IF(MM.EQ.1) GO TO 231
  229 T1=(RAND-VALINT(MM-1,J))/(VALINT(MM,J)-VALINT(MM-1,J))
      S=(MM-1)/10.
      SMALL1=.1*T1+S
      GO TO 233
  231 SMALL1=.1*(RAND/VALINT(1,J))
  233 TOTAL1=TOTAL1+SMALL1*(H(J)-L(J))+L(J)
      TOTAL4=TOTAL4+RAND*(H(J)-L(J))+L(J)
  190 CONTINUE
      TOTAL1=TOTAL1+TPUT
      TOTAL4=TOTAL4+TPUT
      SUM1=SUM1+TOTAL1
      SUM4=SUM4+TOTAL4
      SUMSQ1=SUMSQ1+TOTAL1**2.
      SUMSQ4=SUMSQ4+TOTAL4**2.
      COST1=(TOTAL1-XLOW)/RANGE
      J1=COST1*KK+1.
      IF(J1.EQ.KK+1) J1=KK
      F1(J1)=F1(J1)+1.
      COST4=(TOTAL4-XLOW)/RANGE
      J4=COST4*KK+1.
      IF(J4.EQ.KK+1) J4=KK
      F4(J4)=F4(J4)+1.
      TOTAL1=0.
      TOTAL4=0.
  180 CONTINUE
$
$ COMPUTE STATISTICS
$
      IC=90
      DO 235 I=1,KK
      P1(I)=FLOAT(F1(I))/M
      P2(I)=FLOAT(F2(I))/M
      P4(I)=FLOAT(F4(I))/M
  235 CONTINUE
      CALL STAT(SUM1,M,KK,P1,W,SUMSQ1,MEAN1,MODE1,VAR1,STD1)
      CALL CI(MEAN1,IC,P1,KK,W,WIDTH,L1,U1)
      CALL CDF(P1,MEAN1,KK,W,C1,PX1)
      CALL STAT(SUM2,M,KK,P2,W,SUMSQ2,MEAN2,MODE2,VAR2,STD2)
      CALL CI(MEAN2,IC,P2,KK,W,WIDTH,L2,U2)
      CALL CDF(P2,MEAN2,KK,W,C2,PX2)
      CALL STAT(SUM4,M,KK,P4,W,SUMSQ4,MEAN4,MODE4,VAR4,STD4)
      CALL CI(MEAN4,IC,P4,KK,W,WIDTH,L4,U4)
      CALL CDF(P4,MEAN4,KK,W,C4,PX4)
      IF(N1.EQ.0) GO TO 252
      DO 250 I=1,N1
      MEAN(I)=((A(I)*H(I)+B(I)*L(I)+H(I)+L(I))/(A(I)+B(I)+2.))
      VAR(I)=((A(I)+1.)*(B(I)+1.)*((H(I)-L(I))**2.))/(((A(I)+B(I)+2.)**2.)*(A(I)+B(I)+3.))
  250 CONTINUE
```

```
252 CONTINUE
MSUM=TPUT
VSUM=0.
DO 255 I=1,NTWO
MSUM=MSUM+MEAN(I)
VSUM=VSUM+VAR(I)
255 CONTINUE
%
MEAN3=(XHIGH+XLOW)/2.
VAR3=(1./12.)*RANGE**2.
STD3=SQRT(VAR3)
UNI=1./KK
DO 290 I=1,KK
P3(I)=UNI
290 CONTINUE
CALL CI(MEAN3,IC,P3,KK,W,WIDTH,L3,U3)
CALL CDF(P3,MEAN3,KK,W,C3,PX3)
%
% PRINT INPUT
%
WRITE(II,295)
295 FORMAT(1H )
WRITE(II,300)
300 FORMAT(/)
WRITE(II,310)
310 FORMAT(//)
WRITE(II,320)
320 FORMAT(////)
WRITE(II,330) TITLE
330 FORMAT(1H ,55X,15A4)
WRITE(II,320)
WRITE(II,335)
335 FORMAT(1H ,55X,11HI N P U T S)
WRITE(II,320)
WRITE(II,390) M
390 FORMAT(1H ,15HOBSERVATIONS = ,I5)
WRITE(II,400) ISD2
400 FORMAT(1H ,7HSEED = ,5X,I8)
WRITE(II,310)
WRITE(II,340)
340 FORMAT(1H ,9HBETA VBLE,7X,6HLOWEST,6X,4HMODE,5X,7HHIGHEST,$
5X,7HU COEFF,5X,5HNMODE,5X,5HALPHA,6X,4HBETA,7X,4HMEAN,5X,8HVARIANCE)
WRITE(II,295)
IF(N1.EQ.0) GO TO 351
WRITE(II,350) (I,L(I),MODE(I),H(I),U(I),NM(I),A(I),B(I),MEAN(I),VAR(I),I=1,N1)
350 FORMAT(1H ,3X,I2,8X,F9.1,1X,F9.1,3X,F9.1,6X,F4.2,7X,F4.2,$
6X,F5.1,5X,F5.1,2X,F9.1,1X,F12.1)
IF(N2.EQ.0) GO TO 352
WRITE(II,300)
351 CONTINUE
WRITE(II,353)
353 FORMAT(1H ,10HOTHER VBLE)
WRITE(II,354) (I,L(I),MODE(I),H(I),MEAN(I),VAR(I),I=NONE,NTWO)
354 FORMAT(1H ,3X,I2,8X,F9.1,1X,F9.1,3X,F9.1,44X,F9.1,1X,F12.1)
352 CONTINUE
WRITE(II,295)
WRITE(II,355) TPUT,TPUT,TPUT
355 FORMAT(1H ,10HTHROUGHPUT,3X,F9.1,13X,F9.1,44X,F9.1)
WRITE(II,300)
WRITE(II,360) XLOW,XHIGH,MSUM,VSUM
360 FORMAT(1H ,8H**SUMS**,5X,F9.1,13X,F9.1,44X,F9.1,3X,F10.1)
IF(N2.EQ.0) GO TO 410
WRITE(II,300)
WRITE(II,320)
WRITE(II,370)
370 FORMAT(1H ,10HOTHER VBLE,50X,3HCDF)
WRITE(II,300)
WRITE(II,380) (I,(VALINT(J,I),J=1,10),I=NONE,NTWO)
380 FORMAT(1H ,3X,I2,20X,10(F6.4,3X))
WRITE(II,320)
%
% PRINT OUTPUT
%
410 CONTINUE
WRITE(II,320)
```

```
      WRITE(II,490)
      490 FORMAT(1H ,55X,11HO U T P U T)
      WRITE(II,320)
      WRITE(II,500)
      500 FORMAT(1H ,8HINTERVAL,10X,5HRANGE,6X,22HINDEPENDENT BETA/OTHER,3X,*
      20HDEPENDENT BETA/OTHER,3X,13HTOTAL UNIFORM,6X,19HINDEPENDENT UNIFORM)
      WRITE(II,295)
      WRITE(II,505)
      505 FORMAT(1H ,36X,3HPDF,4X,3HCDF,12X,3HPDF,4X,3HCDF,10X,*
      3HPDF,4X,3HCDF,12X,3HPDF,4X,3HCDF)
      WRITE(II,300)
      WRITE(II,510) (I,W(I),W(I+1),P1(I),C1(I),P2(I),C2(I),P3(I),C3(I),P4(I),C4(I),I=1,KK)
      510 FORMAT(1H ,2X,I2,4X,F9.1,3H - ,F9.1,5X,F6.4,1X,F6.4,9X,F6.4,1X,F6.4,7X,*
      F6.4,2X,F6.4,8X,F6.4,2X,F6.4)
      WRITE(II,310)
      WRITE(II,520) MEAN1,MEAN2,MEAN3,MEAN4
      520 FORMAT(1H ,8H**MEAN**,27X,F9.1,13X,F9.1,12X,F9.1,13X,F9.1)
      WRITE(II,530) MODE1,MODE2,MODE4
      530 FORMAT(1H ,8H**MODE**,27X,F9.1,13X,F9.1,34X,F9.1)
      WRITE(II,540) VAR1,VAR2,VAR3,VAR4
      540 FORMAT(1H ,12H**VARIANCE**,20X,F12.1,10X,F12.1,9X,F12.1,10X,F12.1)
      WRITE(II,545) STD1,STD2,STD3,STD4
      545 FORMAT(1H ,17H**STD DEVIATION**,17X,F10.1,12X,F10.1,11X,F10.1,12X,F10.1)
      WRITE(II,295)
      620 CONTINUE
      WRITE(II,550) IC,L1,L2,L3,L4,U1,U2,U3,U4
      550 FORMAT(1H ,2H**,I2,23H% CONFIDENCE INTERVAL**,8X,F9.1,1H,,12X,F9.1,1H,,*
      11X,F9.1,1H,,12X,F9.1,1H,/36X,F9.1,13X,F9.1,12X,F9.1,13X,F9.1)
      IF(IC.NE.90) GO TO 570
      WRITE(II,295)
      WRITE(II,560) PX1,PX2,PX3,PX4
      560 FORMAT(1H ,20H**PROB EXCEED MEAN**,22X,F3.2,19X,F3.2,18X,F3.2,19X,F3.2)
      WRITE(II,300)
      570 CONTINUE
      WRITE(II,310)
      WRITE(II,600)
      600 FORMAT(1H ,28HANOTHER CONFIDENCE INTERVAL?)
      READ(JJ,*) IC
      IF(IC.EQ.0) GO TO 610
      CALL CI(MEAN1,IC,P1,KK,W,WIDTH,L1,U1)
      CALL CI(MEAN2,IC,P2,KK,W,WIDTH,L2,U2)
      CALL CI(MEAN3,IC,P3,KK,W,WIDTH,L3,U3)
      CALL CI(MEAN4,IC,P4,KK,W,WIDTH,L4,U4)
      GO TO 620
      610 CONTINUE
      WRITE(II,310)
      WRITE(II,622)
      622 FORMAT(1H ,23HPROB EXCEED SOME VALUE?)
      READ(JJ,*) Z
      IF(Z.EQ.0.) GO TO 623
      CALL CDF(P1,Z,KK,W,C1,PX1)
      CALL CDF(P2,Z,KK,W,C2,PX2)
      CALL CDF(P3,Z,KK,W,C3,PX3)
      CALL CDF(P4,Z,KK,W,C4,PX4)
      WRITE(II,624) Z,PX1,PX2,PX3,PX4
      624 FORMAT(1H ,13H**PROB EXCEED,F10.1,2H**,17X,F3.2,19X,F3.2,18X,F3.2,19X,F3.2)
      GO TO 610
      623 CONTINUE
      WRITE(II,310)
      WRITE(II,630)
      630 FORMAT(1H ,24HADDITIONAL OBSERVATIONS?)
      READ(JJ,*) MSTAR
      IF(MSTAR.EQ.0) GO TO 640
      M=M+MSTAR
      GO TO 160
      640 CONTINUE
      WRITE(II,310)
      WRITE(II,650)
      650 FORMAT(1H ,13HANOTHER SEED?)
      READ(JJ,*) KSD
      IF(KSD.EQ.0) STOP
      KSD2=KSD
      WRITE(II,77)
      READ(JJ,*) M
      GO TO 140
```

```
SUBROUTINE CI(MEAN,IC,P,KK,W,WIDTH,LBD,UBD)
REAL W(1),MEAN,P(1),LWT1,LPROB,LBD,LWT2
DO 250 I=2,KK
J=I
IF(W(I).GE.MEAN) GO TO 260
250 CONTINUE
260 MM=0
CON=.5*(IC/100.)
LWT1=(MEAN-W(J-1))/(W(J)-W(J-1))
270 LPROB=LWT1*P(J-1)
K=1
IF(LPROB.GE.CON) GO TO 290
J1=J-1
DO 280 I=2,J1
K=I
LPROB=LPROB+P(J-I)
IF(LPROB.GE.CON) GO TO 290
280 CONTINUE
CON=2.*CON-LPROB
LBD=W(1)
GO TO 300
290 LWT2=(LPROB-CON)/P(J-K)
LBD=W(J-K)+LWT2*WIDTH
IF(MM.EQ.1) GO TO 3 0
%
300 RWT1=1.-LWT1
RPROB=RWT1*P(J-1)
LL=1
IF(RPROB.GE.CON) GO TO 320
J2=KK-J+2
DO 310 I=2,J2
LL=I
RPROB=RPROB+P(J+I-2)
IF(RPROB.GE.CON) GO TO 320
310 CONTINUE
CON=2.*CON-RPROB
UBD=W(KK+1)
MM=1
GO TO 270
320 RWT2=(RPROB-CON)/P(J+LL-2)
UBD=W(J+LL-1)-RWT2*WIDTH
330 RETURN
```

```
SUBROUTINE CDF(P,Z,KK,W,C,PROBX)
REAL P(1),W(1),C(1)
C(1)=P(1)
DO 10 I=2,KK
C(I)=C(I-1)+P(I)
10 CONTINUE
N=KK+1
DO 20 I=2,N
J=I
IF(W(I).GT.Z) GO TO 30
20 CONTINUE
30 A=(Z-W(J-1))/(W(J)-W(J-1))
PROBX=(1.-A)*P(J-1)
DO 40 I=J,KK
PROBX=PROBX+P(I)
40 CONTINUE
RETURN
```

```
      SUBROUTINE DQG32(XL,XU,Y,ALPHA,BETA)
      DOUBLE PRECISION XL,XU,Y,A,B,C,FCT,ALPHA,BETA
      FCT(X)=X**ALPHA*(1.-X)**BETA
      A=.5D0*(XL+XU)
      B=XU-XL
      C=.49863193092474078D0*B
      Y=(.35093050047350483D-2)*(FCT(A+C)+FCT(A-C))
      C=.49280575 7726341700*B
      Y=Y+(.81371973654528350D-2)*(FCT(A+C)+FCT(A-C))
      C=.48238112779375322D0*B
      Y=Y+(.12696032654631030D-1)*(FCT(A+C)+FCT(A-C))
      C=.46745303796886984D0*B
      Y=Y+(.17136931456510717D-1)*(FCT(A+C)+FCT(A-C))
      C=.44816057788302606D0*B
      Y=Y+(.21417949011113340D-1)*(FCT(A+C)+FCT(A-C))
      C=.42468380686628499D0*B
      Y=Y+(.2549 02963118 08 D-1)*(FCT(A+C)+FCT(A-C))
      C=.39724189798397120D0*B
      Y=Y+(.29342046739267774D-1)*(FCT(A+C)+FCT(A-C))
      C=.36609105937014484D0*B
      Y=Y+(.32911111388180923D-1)*(FCT(A+C)+FCT(A-C))
      C=.33152213346510760D0*B
      Y=Y+(.36172897054-24253D-1)*(FCT(A+C)+FCT(A-C))
      C=.29385787862038116D0*B
      Y=Y+(.39096947893535153D-1)*(FCT(A+C)+FCT(A-C))
      C=.25344995446611470D0*B
      Y=Y+(.41655962113473378D-1)*(FCT(A+C)+FCT(A-C))
      C=.21067563 0653176D0*B
      Y=Y+(.43826046502201906D-1)*(FCT(A+C)+FCT(A-C))
      C=.16593430114106382D0*B
      Y=Y+(.45586939347881942D-1)*(FCT(A+C)+FCT(A-C))
      C=.11964368112606854D0*B
      Y=Y+(.46922199540402283D-1)*(FCT(A+C)+FCT(A-C))
      C=.72235980791398250D-1*B
      Y=Y+(.47819360039637430D-1)*(FCT(A+C)+FCT(A-C))
      C=.24153832843869158D-1*B
      Y=B*(Y+(.48270044257363900D-1)*(FCT(A+C)+FCT(A-C)))
      RETURN
      END

      SUBROUTINE STAT(SUM,M,KK,P,W,SUMSQ,MEAN,MODE,VAR,STD)
      REAL MEAN,MODE,P(1),W(1)
      MEAN=SUM/M
      MODE=0.
      DO 10 I=1,KK
      IF(P(I).GT.MODE) MODE=P(I)
      IF(P(I).EQ.MODE) IMODE=I
   10 CONTINUE
      MODE=.5*(W(IMODE)+W(IMODE+1))
      VAR=(SUMSQ-M*MEAN**2.)/(M-1)
      STD=SQRT(VAR)
      RETURN
```

Optimal Sample Size

In most sampling procedures the larger the sample the closer the sample distribution approximates the true distribution. But larger samples are more expensive to generate than smaller ones. The simple experiments described below represent an attempt to determine an optimal sample size for Program SPET -- the smallest sample size that will ensure "reasonable" accuracy in the sampling procedure.

A statistic called the K statistic in this study is used in the search for the optimal sample size. It is defined as

$$K = \sum_{i=1}^{N} (o_i - e_i)^2$$

where N = the number of class intervals

o_i = the number of observations occurring in the ith interval ÷ M.

M = the total number of observations

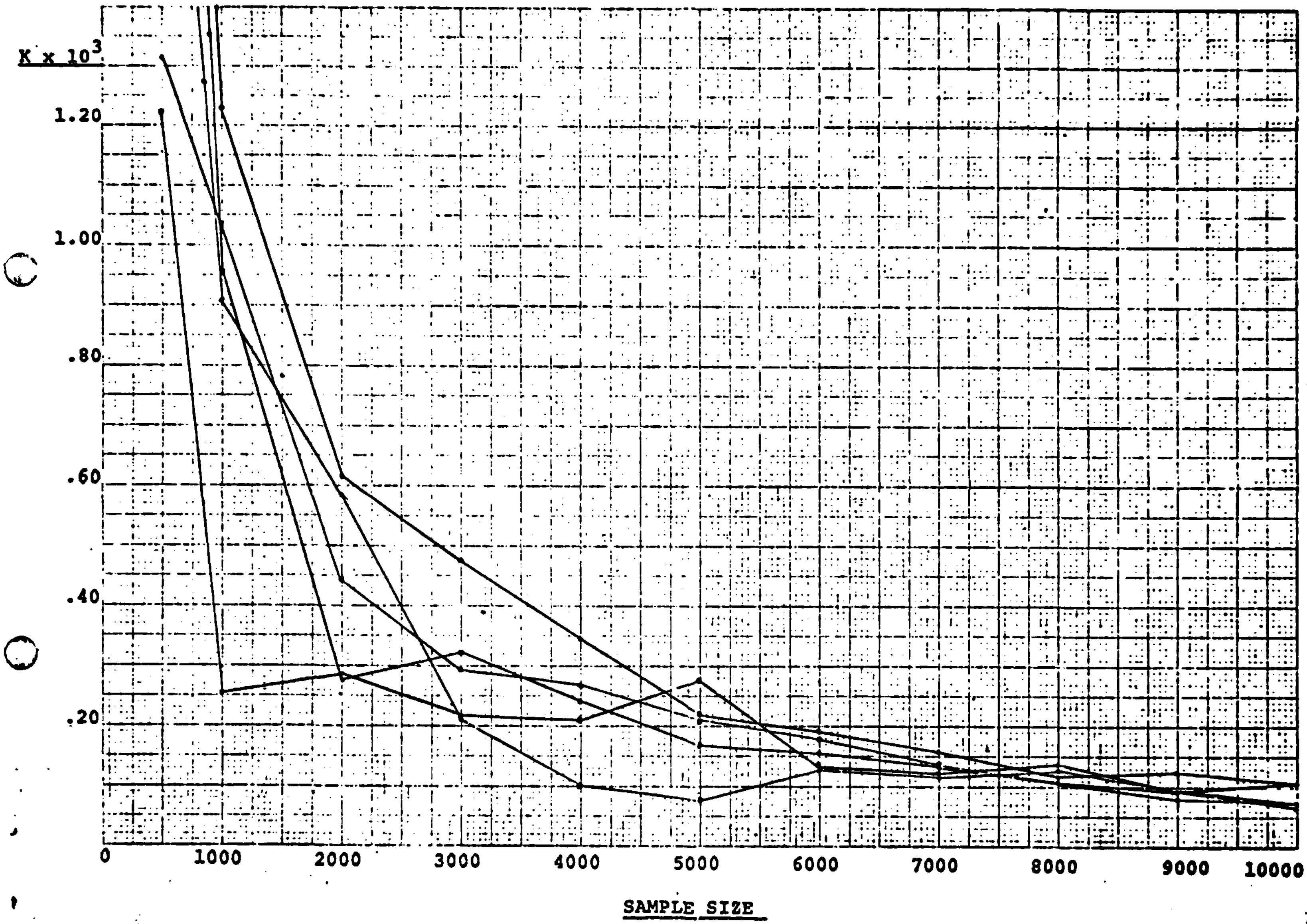
e_i = the number of observations expected in the ith interval (given that the process generating the observations is following a particular statistical distribution) ÷ M.

The K statistic is the sum of squared deviations of the observed probabilities from the expected probabilities of each class interval. As the size of a randomly drawn sample is increased, the K statistic decreases in value until $\lim_{M \to \infty} K = 0$.

The first experiment consists of using five randomly selected seeds with the uniform random number generator used by Program SPET to generate five sequences of K statistics. Each sequence contains a K statistic for sample sizes 500, 1000, 2000, ..., 10000. These K statistics are plotted in Figure C-1 on page C-16. Note how the sequences converge at sample size 6000. It appears that this might be the optimal sample size. Can one expect a sample size of 6000 to ensure "reasonable" accuracy in the sampling procedure?

FIGURE C-1

FIVE SEQUENCES OF K STATISTICS AS A FUNCTION OF SAMPLE SIZE



The accuracy of the Monte Carlo sampling procedure used in Program SPET, for purposes of this inquiry, is measured in terms of the percent deviation of certain statistics from their true values. The second experiment is an attempt to measure the accuracy of Program SPET for various sample sizes. It consists of using the same five seeds selected in the first experiment to draw five sequences of samples from a uniform distribution. Each sequence contains samples of size 500, 1000, 2000, 3000, 6000, and 9000. The mean and the lower and upper confidence limits of the 90% confidence interval are noted from the output of Program SPET and the percent deviation of these statistics from their population values is computed. Then the maximum of the absolute value of the deviations is selected for each statistic in every sample size and plotted in Figure C-2 on page C-18. Note that the rate of decrease in the error (maximum percent deviation) of these statistics is rapid in the range of the sample sizes 500 to 2000, slowing somewhat after sample size 2000.

Consider the error in sample size 2000. Would the expectation of a deviation of at most .21 percent in the mean and 2.44 and 1.81 percent in the lower and upper limits of the confidence interval respectively, be "reasonable?" The authors would answer affirmatively. Reasonableness is subjective. It is felt that the accuracy of sample size 6000 is not enough better than that of sample size 2000 to warrant incurring the increased cost of generating an additional 4000 observations.

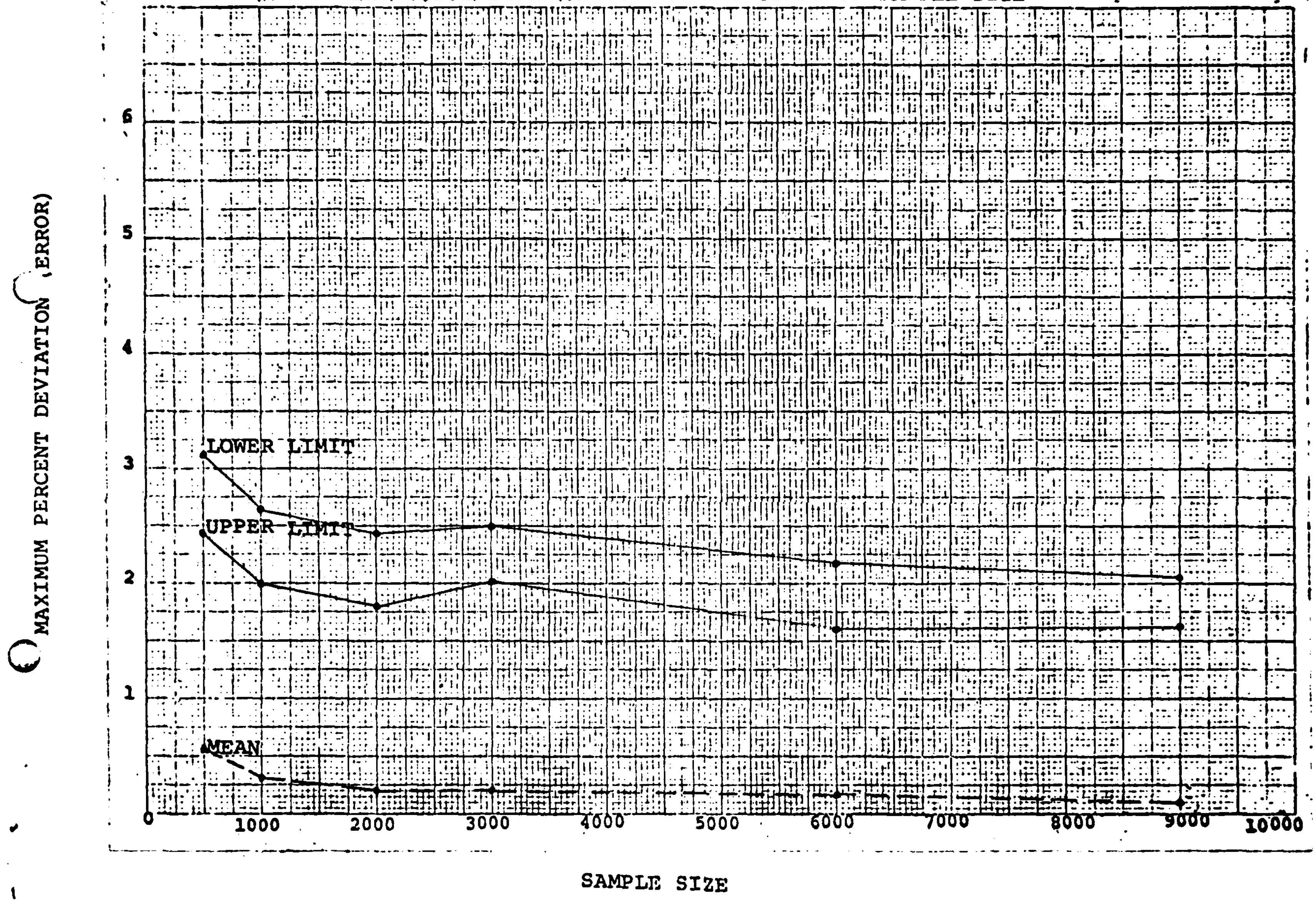
Much greater confidence could be placed in these tentative observations if, instead of five sequences, 30, 40 or more sequences had been generated.[4/] But even the results of the five sequences permit a more confident choice of sample size than no experimentation at all.

4/ *Along with a greater number of sequences one might have repeated experiment two using one or two representative beta distributions in addition to the uniform distribution used above.*

FIGURE C-2

MAXIMUM PERCENT DEVIATION (ERROR) OF CERTAIN STATISTICS

FROM THEIR POPULATION VALUES AS A FUNCTION OF SAMPLE SIZE



REFERENCES

1. Convolution of Inverse Beta Distributions by a Sampling Technique. Bethesda, MD: Mathematica, Inc., 1971.

2. Dienemann, Paul F., Estimating Uncertainty Using Monte Carlo Techniques. Santa Monica, CA: The RAND Corp., RM-4854-PR, 1966.

3. Hillier, F. S. and Lieberman, G. J., Introduction to Operations Research. San Francisco, CA: Holden-Day, Inc., 1967.

4. Kammerer, J. T., ASW Force Level Study - Equipment Readiness: Models, Computer Simulation and Results. Washington, DC: Office of the Chief of Naval Operations, 1968.

5. Larson, Harold J., Introduction to Probability Theory and Statistical Inference. New York: John Wiley and Sons, Inc., 1969.

6. Lindgren, Bernard W., Statistical Theory. 2nd ed. New York: The MacMillian Co., 1968.

7. MacCrimmon, K. R. and Ryavec, C. A., An Analytical Study of the PERT Assumptions. Santa Monica, CA: The RAND Corp., RM-3408-PR, 1962.

8. Schaefer, Donald F., et. al., A Monte Carlo Simulation Approach to Cost-Uncertainty Analysis. McLean, VA: Research Analysis Corp., RAC-TP-349, 1969.

9. Sobel, S., A Computerized Technique to Express Uncertainty in Advanced System Cost Estimates. Bedford, MA: The MITRE Corp., TM-3728, 1963.

10. Sutherland, William H., Fundamentals of Cost Uncertainty Analysis. McLean, VA: Research Analysis Corp., RAC-CR-4, 1971.

11. ______________________, A Method for Combining Asymmetric Three-Value Predictions of Time or Cost. McLean, VA: Research Analysis Corp., RAC-P-65, 1972.